

Universidade de Lisboa
Faculdade de Letras
Departamento de Linguística Geral e Românica



**Complexidade textual e progressão escolar em dois registos:
um estudo de correlação baseado em um *corpus quasi-longitudinal***

Mário Gleisse das Chagas Martins

Orientadores: Professor Doutor Rui Pedro Ribeiro Marques
Professor Doutor Telmo Lopes Mória

Tese especialmente elaborada para a obtenção do grau de doutor em
Linguística, na especialidade Linguística Educacional.

2016

Universidade de Lisboa
Faculdade de Letras
Departamento de Linguística Geral e Românica



**Complexidade textual e progressão escolar em dois registos:
um estudo de correlação baseado em um *corpus quasi-longitudinal***

Mário Gleisse das Chagas Martins

Orientadores: Professor Doutor Rui Pedro Ribeiro Marques
Professor Doutor Telmo Lopes Mória

Tese especialmente elaborada para a obtenção do grau de doutor em Linguística, na especialidade Linguística Educacional.

Júri

Presidente:

Doutora **Maria Inês Pedrosa da Silva Duarte**, Professora Catedrática e Membro do Conselho Científico da Faculdade de Letras da Universidade de Lisboa.

Vogais:

Doutora **Ana Cristina Macário Lopes**, Professora Associada com Agregação da Faculdade de Letras da Universidade de Coimbra;

Doutora **Ana Luísa da Piedade Melro Blazer Gaspar Costa**, Professora Adjunta da Escola Superior de Educação do Instituto Politécnico de Setúbal;

Doutora **Maria Amália Pereira Mendes**, Investigadora Auxiliar do Centro de Linguística da Universidade de Lisboa;

Doutora **Maria Inês Pedrosa da Silva Duarte**, Professora Catedrática da Faculdade de Letras da Universidade de Lisboa;

Doutora **Anabela Proença Leitão Gonçalves**, Professora Associada da Faculdade de Letras da Universidade de Lisboa;

Doutora **Maria Armanda Martins da Costa**, Professora Auxiliar da Faculdade de Letras da Universidade de Lisboa;

Doutor **Rui Pedro Ribeiro Marques**, Professor Auxiliar da Faculdade de Letras da Universidade de Lisboa, orientador.

Esta investigação foi financiada pela Fundação para a Ciência e a Tecnologia, sob o número de bolsa BD/60601/2009.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

Como sempre, para ti, Tonká...
de cujas mãos o Tejo nunca escorreu.

Agradeço...

À **Fundação para a Ciência e Tecnologia**, que apoiou financeiramente a realização desta investigação.

Aos professores **Rui Marques** e **Telmo Móia**, que confiaram em mim, sendo sempre pacientes e participativos e estando sempre disponíveis para longos ‘cafés’. Tudo isto sob o signo da gentileza. A atuação destes professores enobrece o significado da palavra “orientação”.

À Professora **Amália Mendes**, que, além de sugestões muito ricas sobre questões teóricas e metodológicas, contribuiu enormemente para a anotação do *corpus*.

Aos professores **Márcia Nogueira**, **Michael O’Donnell**, **Anabela Gonçalves** e **Armanda Costa**, que me deram sugestões teóricas e metodológicas de muita valia.

Ao meu amigo **Lachlan Mackenzie**, que, na hora mais escura, me mostrou o caminho iluminado.

Ao meu amigo **William Martinez**, que sempre foi um bom ouvinte/leitor e um bom professor de estatística.

À minha amiga **Mafalda Mendes**, com quem tive excelentes discussões sobre o desenvolvimento linguístico de crianças e adolescentes.

À minha amiga **Susana Castro**, que sempre me ajudou em tudo, inclusive nos momentos em que uns copos eram necessários.

A todos os colegas que, de um modo ou de outro, contribuíram para que esta investigação se realizasse, em especial ao **Sóstenes Rego**, à **Sílvia Barbosa**, ao **Pedro Morais** e ao **Paulo Barroso**.

Aos **alunos** e aos seus **encarregados de educação**, aos **professores** e às **escolas**, por terem participado deste projeto.

À minha **mãe** e às minhas **irmãs**, que, perto ou longe, estão sempre ao meu lado.

Resumo

Esta investigação tem por objetivo apresentar um estudo de correlação entre indicadores de complexidade textual e progressão escolar. A complexidade textual é representada por indicadores pertencentes às dimensões lexical e sintática da língua, referenciados pela literatura sobre o desenvolvimento linguístico, em particular Hunt (1965, 1970), Wolfe-Quintero *et al.* (1998) e Berman (2007), como indicadores dos graus de maturidade do uso que alunos em diferentes níveis da instrução formal fazem da língua ao longo da progressão nos anos escolares. Para atingir o objetivo proposto, foram selecionados como participantes 122 alunos monolíngues de português europeu inscritos no 5.º, no 7.º e no 10.º ano de escolaridade, representando cada um dos três ciclos da escolaridade básica em Portugal. Cada aluno produziu um texto de registo narrativo e um outro de registo argumentativo, o que resultou na obtenção de 244 textos. Todos os textos foram produzidos a partir dos mesmos estímulos e sob condições semelhantes, permitindo a compilação de um *corpus* do tipo *quasi*-longitudinal (Granger 2008). Aos textos obtidos foram aplicadas automaticamente etiquetas morfossintáticas, tendo, além disso, sido feitas manualmente a segmentação e a anotação em orações e unidades-t. Ao considerar, em linha com Wilkinson *et al.* (1980: 2), que, na escola, o desenvolvimento linguístico naturalmente acontece, as hipóteses de trabalho podem ser resumidas na expectativa de que há uma correlação positiva entre os indicadores de complexidade textual e a progressão escolar, correlação esta examinada sob os constrangimentos específicos de cada um dos registos e das tarefas de escrita. Os resultados obtidos indicam que, quanto à complexidade lexical, há uma correlação positiva apenas entre a diversidade e a progressão escolar, não sendo detetada qualquer correlação entre a densidade, tomada globalmente, e a progressão escolar. No entanto, a apreciação dos itens por especialidade (nomes, verbos, adjetivos e advérbios terminados em *-mente*) revela que há uma correlação positiva entre a frequência de nomes e a progressão e, ao contrário, há uma correlação negativa entre a frequência de verbos e a progressão. Quanto à complexidade sintática, em ambos os registos, a correlação é dependente do tipo de oração subordinada. As orações completivas não mantêm correlação alguma com a progressão; as adverbiais mantêm uma correlação negativa com a progressão e as orações relativas e gradativas mantêm uma correlação positiva com a progressão escolar. Complementarmente aos estudos de correlação, apresentam-se breves estudos de caso sobre os usos que os alunos fazem da língua, justificados em particular por frequências expressivas estatisticamente.

Palavras-chave: complexidade textual, progressão escolar, desenvolvimento linguístico em idade escolar, registo.

Abstract

This correlational study aims to investigate the relationship between school grading and textual complexity measures, devised on lexical and syntactic language dimensions. In order to describe the different stages of children and adolescent language development along school years, these measures are employed in literature on language development, namely in Hunt (1965, 1970), Wolfe-Quintero *et al.* (1998) and Berman (2007). To achieve the aim of this study, 122 basic schooling students were recruited from years 5, 7 and 10 of four different public schools. All of the students are native speakers of European Portuguese. Under the same conditions and from the same *stimuli*, each student wrote one narrative and one argumentative text, resulting in a *quasi*-longitudinal *corpus* of 244 texts. A complete lemmatization and POS tagging were provided for this data. As for clause and t-unit segmentation, the *corpus* was manually tagged. In accordance with Wilkinson *et al.*'s assertion (1980: 2) that, in school, “development obviously takes place”, the main hypothesis behind this study is that there is a positive correlation between textual complexity metrics and school grading, examined under specific constraints of each register and writing tasks. As for lexical complexity, the findings reveal a positive correlation between diversity and school grading and no correlation between density and school grading, although the analysis of specific word classes (nouns, verbs, adjectives and adverbs in –mente) points, on one hand, to a positive correlation between nouns and school grading and, on the other hand, to a negative correlation between verbs and school grading. As for syntactic complexity, the findings suggest that in both narrative and argumentative registers the type of correlation is dependent on the type of subordination. There is no correlation between completive clauses and school grading; adverbial clauses keep a negative correlation with school grading. Only relative and gradative clauses maintain a positive correlation with school grading. In addition to correlational study, case studies on student's language uses were undertaken, justified mostly by significant frequencies.

Keywords: textual complexity, school grading, school-age language development, register.

ÍNDICE

Resumo.....	i
Abstract.....	iii
Lista de Tabelas	vii
Lista de Gráficos	viii
1. Apresentação do tema e síntese da proposta de trabalho	1
2. Definições operacionais	11
2.1. Complexidade	11
2.2. Texto	14
2.3. Contexto (e cotexto).....	20
2.4. Registo	27
3. Procedimentos metodológicos gerais	30
3.1. Caracterização do estudo.....	30
3.2. Escolas e participantes	31
3.3. Compilação do corpus.....	34
3.3.1. Transcrição e armazenamento.....	35
3.3.2. Ferramentas de anotação e consulta (<i>query</i>)	37
4. Complexidade lexical.....	41
4.1. Introdução	41
4.2. Diversidade.....	43
4.3. Densidade	46
4.4. Procedimentos metodológicos específicos	49
4.5. Estudo de correlação	51
4.6. Discussão	57
4.7. Síntese	62
5. Complexidade sintática	65
5.1. Introdução	65
5.2. Extensão	69
5.3. Variação	71
5.4. Procedimento metodológicos específicos	74
5.5. Estudo de correlação	78

5.6.	Discussão	87
5.7.	Síntese	92
6.	Estudos de casos.....	95
6.1.	Estatuto lexical do verbo ‘ir’	98
6.2.	Expressão do tempo adjunto e argumental.....	100
6.3.	Influência extralinguística na variação vocabular	103
6.4.	Alternância no emprego de verbos epistêmicos.....	103
6.5.	Construções com o verbo ‘achar’	104
6.6.	Incidência de pronomes pessoais e demonstrativos.....	105
6.7.	Orações relativas	111
6.7.1.	Diversidade de uso e função sintática de pronomes relativos	115
6.7.2.	(In)Adequação no emprego do pronome relativo “que”	120
6.8.	Orações adverbiais	124
6.8.1.	Localização temporal entre situações e posição das orações temporais.....	127
6.8.2.	Diversidade de uso dos conectores das orações causais.....	135
6.9.	Síntese	141
7.	Considerações finais	144
7.1.	Implicações pedagógicas	144
7.2.	Limitações	151
7.3.	Desdobramentos desta investigação.....	152
8.	Referências bibliográficas	156
9.	ANEXOS.....	168
9.1.	Protocolo de cooperação	168
9.2.	Termo de consentimento.....	169
9.3.	Questionário para a identificação da L1 dos participantes.....	170
9.4.	Variáveis de investigação	171
9.5.	Valores absolutos	172
9.6.	Orientações gerais de segmentação e anotação de orações.....	173
9.7.	Tipos e subtipos de oração subordinada	176

Lista de Tabelas

Tabela 1: Frequência média (e desvio-padrão), em números absolutos, de ocorrência de palavras totais (<i>tok</i>) e palavras diferentes (<i>typ</i>) identificada nos registos narrativo e argumentativo.....	51
Tabela 2: Frequência média (e desvio-padrão), em números absolutos, de itens lexicais (<i>lex</i>) e de itens gramaticais (<i>gram</i>) identificados nos registos narrativo e argumentativo.....	53
Tabela 3: Valores absolutos das ocorrências de unidades-t (<i>unt</i>), unidades-t simples (<i>sim</i>), unidades-t multioracionais (<i>mul</i>), orações (<i>orc</i>), orações subordinadas (<i>sub</i>) e palavras (<i>tok</i>) nos registos narrativo e argumentativo.....	77
Tabela 4: Valores médios de ocorrência, por texto, de palavras (<i>Tok</i>), de orações (<i>Orc</i>), de orações subordinadas (<i>Sub</i>) e de unidades-t (<i>Unt</i>) identificados nos registos narrativo e argumentativo.....	78
Tabela 5: Valores médios de ocorrência, por texto, de orações completivas (<i>Com</i>), adverbiais (<i>Adv</i>), relativas (<i>Rel</i>) e gradativas (<i>Gra</i>) identificados nos registos narrativo e argumentativo.....	83
Tabela 6: Lista das dez palavras lexicais mais frequentes (<i>tokens</i>) no registo narrativo ao longo da progressão escolar.....	96
Tabela 7: Lista das dez palavras lexicais mais frequentes no registo argumentativo ao longo da progressão escolar.....	97
Tabela 8: Lista das dez formas mais frequentes (<i>types</i>) do verbo ‘ir’ no registo narrativo ao longo da progressão escolar.....	98

Lista de Gráficos

Gráfico 1: Distribuição dos valores médios da medida <i>D</i> , em números absolutos, nos registos narrativo e argumentativo ao longo dos anos escolares.....	52
Gráfico 2: Distribuição dos valores médios da medida <i>Z</i> , em valores percentuais, nos registos narrativo e argumentativo ao longo dos anos escolares.....	54
Gráfico 3: Distribuição dos valores médios de ocorrências, em percentual, de nomes comuns, verbos plenos, adjetivos e advérbios terminados em <i>-mente</i> no registo narrativo ao longo dos anos escolares.....	55
Gráfico 4: Distribuição dos valores médios de ocorrências, em percentual, de nomes comuns, verbos plenos, adjetivos e advérbios terminados em <i>-mente</i> no registo argumentativo ao longo dos anos escolares.....	56
Gráfico 5: Frequência média de palavras por oração (<i>Tok/Orc</i>), de palavras por unidade- <i>t</i> (<i>Tok/Unt</i>) e de orações subordinadas por unidade- <i>t</i> (<i>Sub/Unt</i>) no registo narrativo ao longo dos anos escolares.....	79
Gráfico 6: Frequência média de palavras por oração (<i>Tok/Orc</i>), de palavras por unidade- <i>t</i> (<i>Tok/Unt</i>) e de orações subordinadas por unidade- <i>t</i> (<i>Sub/Unt</i>) no registo argumentativo ao longo dos anos escolares.....	80
Gráfico 7: Valores médios do rácio de orações completivas (<i>Com</i>), orações adverbiais (<i>Adv</i>), orações relativas (<i>Rel</i>) e orações gradativas (<i>Gra</i>) pelo total de orações subordinadas (<i>Sub</i>) identificados nos textos do registo narrativo.	84
Gráfico 8: Valores médios do rácio de orações completivas (<i>Com</i>), orações adverbiais (<i>Adv</i>), orações relativas (<i>Rel</i>) e orações gradativas (<i>Gra</i>) pelo total de orações subordinadas (<i>Sub</i>) identificados nos textos do registo argumentativo.	84
Gráfico 9: Distribuição dos valores médios de ocorrências, em percentual, de construções com o verbo ‘ <i>ir</i> ’ no registo narrativo ao longo dos anos escolares.	99
Gráfico 10: Distribuição dos valores médios de ocorrência, em percentual, de sintagmas nominais e preposicionais com a função de adjuntos temporais construídos com o item ‘ <i>dia</i> ’ no registo narrativo ao longo da progressão escolar.	101
Gráfico 11: Distribuição dos valores médios de ocorrências, em percentual, de construções com predicados verbais temporais no registo narrativo ao longo dos anos escolares.	102

Gráfico 12: Distribuição dos valores médios de ocorrências, em percentual, de verbos modais epistêmicos no registo argumentativo ao longo dos anos escolares.	104
Gráfico 13: Distribuição dos valores médios de ocorrências, em percentual, de construções com o verbo epistêmico ‘achar’ no registo argumentativo ao longo dos anos escolares.....	105
Gráfico 14: Distribuição da frequência normalizada de pronomes anafóricos no registo narrativo ao longo dos anos escolares.....	109
Gráfico 15: Distribuição da frequência normalizada de pronomes anafóricos no registo argumentativo ao longo dos anos escolares.....	110
Gráfico 16: Distribuição dos valores médios de ocorrências, em percentual, de orações relativas nos registos narrativo e argumentativo ao longo dos anos escolares.	112
Gráfico 17: Distribuição dos valores médios de ocorrências, em percentual, de relativas restritivas, apositivas e de frase no registo narrativo ao longo dos anos escolares.	114
Gráfico 18: Distribuição dos valores médios de ocorrências, em percentual, de relativas restritivas, apositivas e de frase no registo argumentativo ao longo dos anos escolares.....	114
Gráfico 19: Distribuição dos valores médios de ocorrências, em percentual, de pronomes relativos no registo narrativo ao longo dos anos escolares.	116
Gráfico 20: Distribuição dos valores médios de ocorrências, em percentual, de pronomes relativos no registo argumentativo ao longo dos anos escolares.	116
Gráfico 21: Distribuição dos valores médios de ocorrências, em percentual, das funções sintáticas exercidas pelo morfema ‘que’ dentro de orações relativas no registo narrativo ao longo dos anos escolares.....	118
Gráfico 22: Distribuição dos valores médios de ocorrências, em percentual, das funções sintáticas exercidas pelo morfema ‘que’ dentro de orações relativas no registo argumentativo ao longo dos anos escolares.....	118
Gráfico 23: Distribuição dos valores médios de ocorrências adequadas, em percentual, do pronome relativo ‘que’ nos registos narrativo e argumentativo ao longo dos anos escolares.....	121
Gráfico 24: Distribuição dos valores médios de ocorrências, em percentual, de orações adverbiais nos registos narrativo e argumentativo ao longo dos anos escolares..	125

Gráfico 25: Distribuição dos valores médios de ocorrências, em percentual, dos tipos de orações subordinadas adverbiais no registo narrativo ao longo dos anos escolares.	126
Gráfico 26: Distribuição dos valores médios de ocorrências, em percentual, dos tipos de orações subordinadas adverbiais no registo argumentativo ao longo dos anos escolares.	126
Gráfico 27: Distribuição dos valores médios de ocorrências, em percentual, de conectores empregues nas orações temporais no registo narrativo ao longo dos anos escolares.	132
Gráfico 28: Distribuição dos valores médios de ocorrências, em percentual, dos tipos de relação entre a localização temporal da oração temporal (anterioridade, posterioridade e sobreposição) e a sua posição (inicial, medial e final) no registo narrativo ao longo dos anos escolares.	133
Gráfico 29: Distribuição dos valores médios de ocorrências, em percentual, dos tipos de conectores empregues nas orações causais no registo argumentativo ao longo dos anos escolares.	138
Gráfico 30: Distribuição dos valores médios de ocorrências, em percentual, dos tipos de conectores empregues nas orações causais no registo narrativo ao longo dos anos escolares.	139

1. Apresentação do tema e síntese da proposta de trabalho

Nesta dissertação, tem-se por objetivo explicitar um estudo de correlação entre indicadores de complexidade textual e progressão escolar, identificando-se potenciais regularidades e tendências no uso que alunos da educação básica fazem da língua escrita. Para tal, um conjunto de indicadores de complexidade textual referenciados pela literatura é observado num *corpus* constituído por textos de registo narrativo e argumentativo escritos por três grupos de alunos monolíngues de português europeu, cada grupo procedente de um dos três ciclos da escolaridade básica.

É um estudo que se enquadra na área de investigação conhecida hoje como “desenvolvimento linguístico em idade escolar” (Berman 2007: 348)¹, em que se examina o processo gradual, por que passam crianças e adolescentes, de familiarização com a língua escrita, em particular em contextos comunicativos diversificados (Ravid e Tolchinsky 2002: 2). A língua escolar, neste sentido, conforma o estágio precursor da língua adulta e escolarizada (Ravid 2004: 337), ao mesmo tempo que é um estágio posterior à aquisição.² Embora esta seja uma área de investigação em ascensão nas últimas quatro décadas, ainda se encontra num estágio bastante incipiente. A lacuna é particularmente evidente no que respeita às pesquisas sobre a modalidade escrita, que continua a ser uma modalidade pouco investigada, principalmente no tocante à língua materna (Sampson 2005; Myhill e Fisher 2010).

¹ Alguns trabalhos, em especial das décadas de 70 e 80 do século passado, como Karmiloff-Smith (1986), sustentam-se num conceito etário e, por isso, optam por usar a expressão “later language development” (desenvolvimento linguístico tardio, como prefere Costa 2010). Quando o desenvolvimento se refere apenas à língua escrita, também é conhecido, em inglês, como “writing development” (Myhill e Fisher 2010: 405). Note-se que o desenvolvimento linguístico em idade escolar é parte de uma área de investigação maior, que é a da aquisição e desenvolvimento em língua materna e em língua segunda/estrangeira (Wray *et al.* 2001).

² Em linha com Wray *et al.* (2001: 40), faz-se aqui a distinção entre o processo de aquisição e o de desenvolvimento. O primeiro refere-se à aprendizagem das estruturas e regras da língua, nas suas dimensões sintática, fonológica, etc.; o segundo, ao uso (e reelaboração) que se faz das estruturas e regras já adquiridas.

Segundo Karmiloff-Smith (1986: 456), a ausência de estudos sobre o desenvolvimento linguístico em idade escolar foi inicialmente motivada pela crença de que uma criança, pelos quatro ou cinco anos de idade, já domina as estruturas morfossintáticas centrais da sua língua nativa e que, no amadurecimento até à vida adulta, pouco ou nada se altera quanto ao seu uso, observando-se apenas mudanças de natureza lexical.

É Carol Chomsky (1969), segundo Fletcher e Garman (1986: 451), quem primeiro contraria esta crença. Na sua experiência psicolinguística “The Acquisition of Syntax in Children From 5 to 10”, que se tornou num marco para os estudos sobre o desenvolvimento da linguagem em idade escolar, a autora mostra que crianças monolingues, aos cinco anos de idade, ainda não dominam plenamente todas as estruturas sintáticas da língua adulta. A partir de então, lança-se o desafio para que novos estudos demonstrem que, após a aquisição de uma língua, tem continuidade um longo curso de desenvolvimento, o qual se estende gradualmente até à vida adulta, quando finalmente estabiliza.³

Após Chomsky (1969), o desenvolvimento linguístico em idade escolar (ou abreviadamente desenvolvimento linguístico) começa a consolidar-se enquanto área de investigação, de que resulta a assunção de que não se deve assumi-lo como apenas um acumular indiscriminado de novas formas linguísticas. Tolchinsky (2004: 234) sustenta que se deve assumi-lo como um movimento de reorganização funcional, em que formas já adquiridas ganham novas funções e, concomitantemente, funções já compreendidas passam a realizar-se com novas formas.

Esta reorganização funcional, na perspetiva de Gillis e Ravid (2009: 232), depende da interação de outros fatores desenvolvimentais, para além do linguístico: o fator cognitivo, o fator afetivo-social e o fator literacia (associado à escolarização). O fator cognitivo, *grosso modo*, refere-se às capacidades de crianças e adolescentes de processar informações e solucionar problemas a partir da integração de diferentes fontes de conhecimento, o que se caracteriza pela capacidade de definir objetivos, procurar e avaliar opções, planear e monitorizar procedimentos ou eliminar inconsistências e redundâncias. O fator afetivo-social assenta na interação social. É por ela que crianças e jovens entendem o conceito de

³ Kemper (1988, cf. Cheung e Kemper 1992: 54) sugere que há uma mudança na complexidade linguística de adultos que se pode relacionar com a idade.

identidade e alteridade e adquirem novos conhecimentos, que, no final, são responsáveis pela reorganização das suas estruturas de conhecimento. Finalmente, a escolarização, cuja meta é a obtenção de níveis ótimos de literacia, é um dos fatores responsáveis pelo desenvolvimento linguístico, o que se pode descrever melhor se incluído no conceito geral de literacia linguística (Ravid e Tolchinsky 2002). Assim, na escola, sistematiza-se o conhecimento do mundo enquanto condição necessária para a interpretação e a produção de textos em códigos semióticos diversificados. Um exemplo é o enriquecimento do repertório lexical, causado por um robusto conjunto de experiências linguísticas, que ocorrem pela exposição continuada aos diversos tipos de discurso, os quais se materializam nos registos constitutivos das práticas sociais nas comunidades letradas, fomentando o uso funcional da língua (Ravid e Tolchinsky 2002: 420). No processo de desenvolvimento da literacia, a escrita ocupa um papel central, como ilustra a definição de literacia em Olson (2006, cf. Berman e Ravid 2009: 92): “a set of social practices that exploit the affordances of writing for particular ends”. Não se ignora, naturalmente, que uma pessoa letrada também tenha de dominar o modo falado da língua, mas é a escrita a plataforma da literacia.

Enquanto parte do processo de consolidação da área de investigação sobre o desenvolvimento linguístico, os contextos de estudo cada vez mais se diversificam, sendo considerados quer o contexto do desenvolvimento de línguas maternas (Strömquist *et al.* 2004 e Johansson 2009, para o sueco; Vila 1991 e Crespo-Allende e Alfaro-Faccio 2010, para o espanhol; Colleta 2004 e Mazur-Palandre e Jisa 2010, para o francês), quer o de línguas segundas/estrangeiras (Wolfe-Quintero *et al.* 1998), quer ainda o desenvolvimento linguístico em ambientes multilingues (Verhoeven e Strömquist 2001). Quanto à língua materna em particular, constitui-se como uma das principais referências o projeto Spencer (Berman e Verhoeven 2002), em que se analisa o desenvolvimento em composições textuais faladas e escritas de crianças e adolescentes em diferentes idades e ciclos escolares, de diferentes países e línguas. Há ainda a assinalar trabalhos interdisciplinares, tais como estudos sobre os graus de correlação entre o desenvolvimento linguístico e as interferências de fatores sócio-culturais e económicos (Schleifer 2003) ou de fatores étnicos (Abadiano 1995) e sobre o desempenho de pessoas com necessidades especiais ou afasias e outros distúrbios de fala (Laws e Gunn 2002).

No que respeita ao português europeu como língua materna, ainda são escassos os trabalhos de carácter longitudinal ou *quasi*-longitudinal⁴ baseados em *corpora* sobre o desenvolvimento linguístico na produção textual de crianças e adolescentes monolíngues ao longo da progressão escolar. Com caracterização próxima, identificaram-se os seguintes trabalhos: Carvalho (1990), que procura compreender a relação entre a evolução sintática e o desenvolvimento intelectual de alunos do 2.º ao 8.º ano; Pereira (2008), que descreve o carácter situado da linguagem escolar, exemplificando-o na seleção anafórica demonstrativa de alunos do 3.º e do 4.º ano; Rodrigues (2008), que analisa um conjunto de indicadores de complexidade em textos narrativos escritos por crianças monolíngues do 1.º ao 4.º ano; e Costa (2010), que examina o uso de conetores de valor concessivo em produções textuais dos alunos do 4.º, 6.º e 9.º anos. Todos estes estudos para o português se enquadram na área de investigação aqui referida.

Estudos desta natureza contribuem para a procura de soluções para um persistente problema relacionado com o desenvolvimento da escrita escolar, bem sintetizado na seguinte afirmação de Wilkinson *et al.* (1980: 2): “Development obviously takes place, but does not take place obviously”. Quer isto dizer que, apesar da certeza de que a escrita de um rapaz de quatorze anos é provavelmente melhor do que a escrita de um rapaz de onze, que, por sua vez, é melhor do que a escrita de um rapaz de sete, ainda não se tem, com precisão, o conhecimento de como esta melhoria efetua. Este desconhecimento tem influência direta nas pedagogias de escrita que se utilizam quer em sala de aula, quer nos programas curriculares e materiais didáticos, sendo tais pedagogias, por vezes, discordantes sobre a melhor forma de ensinar a escrita.

O trabalho que agora se apresenta surge de duas motivações. A primeira é a de contribuir, a par de Carvalho (1990), Pereira (2008), Rodrigues (2008) e Costa (2010), entre outros, para a ampliação do conhecimento sobre os aspetos do desenvolvimento linguístico

⁴ Como referido já no título deste trabalho, esta investigação baseia-se num *corpus* *quasi*-longitudinal (Granger 2008), em que se obtêm, no mesmo tempo e nas mesmas condições, dados que representam diferentes estádios de desenvolvimento do fenómeno que se quer escrutinar, configurando-se, portanto, uma evolução aparente no tempo, simulando um *corpus* longitudinal. Este tópico será mais bem discutido no Capítulo 3 (p. 29).

em idade escolar na realidade portuguesa, tornando possível a construção de um perfil global dos estádios de aquisição e desenvolvimento da língua de crianças e jovens. A segunda motivação, de ordem pedagógica, é a de contribuir para a comunidade escolar com o fornecimento de informações que sejam pertinentes para a sistematização de programas curriculares, para a elaboração de materiais didáticos e também para a reflexão sobre as práticas do ensino de português como língua materna (e, por implicação, de português como língua segunda/estrangeira).

Ainda quanto ao processo de consolidação dos estudos sobre o desenvolvimento linguístico, é importante referir que uma das mais importantes consequências é ter-se um objeto de investigação bem delimitado. Embora sejam estudos comumente multidisciplinares, em que coatuam disciplinas como a psicologia, a neurociência e a pedagogia, é a linguística a sua disciplina central, como destacado por Ravid e Tolchinsky (2002). Ou seja, uma análise sobre o desenvolvimento linguístico é, primordialmente, uma análise linguística. A língua é a sua substância, e todas as suas unidades são candidatas a tópicos de pesquisa. Por exemplo, pode-se analisar a evolução no uso dos recursos lexicais (Chipere *et al.* 2001; Malvern *et al.* 2004), da derivação morfológica (Ravid 2004), da complementação oracional (Beers e Nagy 2011; Berninger *et al.* 2011) ou do fluxo informacional (Mazur-Palandre e Jisa 2010).

Além de saber como acontece o desenvolvimento linguístico, também é de fundamental importância saber como medi-lo, já que, ao fazê-lo, se tem a possibilidade de distinguir estádios de desenvolvimento e de estipular, com alguma certeza, o quanto as crianças e os adolescentes utilizam do sistema linguístico a que têm acesso. Para tanto, há um conjunto de indicadores já bastante referenciados pela literatura, os quais, de acordo com Beers e Nagy (2011: 185), se manifestam ao longo de dois eixos complementares. Num primeiro eixo, há os indicadores do desenvolvimento que se relacionam com o domínio de um código linguístico mais elaborado e mais extenso, pertencendo à dimensão estrutural ou gramatical da língua. Trata-se dos indicadores de complexidade sintática (ou gramatical) e de complexidade lexical. Num segundo eixo, há os indicadores que configuram o desenvolvimento de capacidades de produção e interpretação de discursos, considerando-se os contextos de uso e, conseqüentemente, a utilização de géneros textuais adequados a tais contextos. São estes os indicadores de complexidade coesiva e de complexidade genológica, ambos pertencentes à dimensão discursiva. Neste estudo, deter-me-ei apenas nos indicadores de complexidade

lexical e sintática e, para designá-los conjuntamente, utilizarei a expressão “indicadores de complexidade textual”.

Na linguística, os estudos que recorrem aos indicadores de complexidade textual seguem uma de duas tendências teóricas: a tendência psicolinguística (ou cognitiva) e a tendência social. A primeira, mais orientada para a aquisição e o desenvolvimento de uma língua segunda ou estrangeira, associa os indicadores de complexidade a níveis de compreensão (ou de legibilidade), colocando-os em paralelo com outros fatores, tais como as capacidades de decodificação (Hinkel 2006), o interesse do aprendiz (Bügel e Buunk 1996) ou o conhecimento cultural (Barry e Lazarte 1998); a tendência social, em grande parte fundamentada em Bernstein (1960, 1971, 1972) e Labov (1972), investiga como o código linguístico pode ser característico de determinados grupos/classes sociais.

Nesta tese, diferentemente dos estudos acima citados, a opção é não me afiliar a uma ou outra destas tendências teóricas, mas a uma tendência orientada para a variação no uso (Wray *et al.* 2001). Deste modo, utilizo os indicadores de complexidade textual como instrumentos para examinar o registo escolar, que, à partida, distingo entre narrativo e argumentativo, mantendo-me conetado com o discurso habitual das aulas de português. Nos moldes da definição de registo proposta por Biber (2006: 11), entendo que o registo escolar é uma variedade de uso da língua definida contextualmente e passível de ser descrita de acordo com as suas características léxico-gramaticais. Portanto, a comparação de indicadores de complexidade aplicados aos dois registos escolares escolhidos pode revelar, ainda em linha com Biber e Conrad (2009: 37), padrões linguísticos próprios da narração e da argumentação escolares, até então transparentes. Disto resulta um estudo descritivo, mas também empírico, pelo que é mandatário enquadrá-lo por alguns princípios de investigação.

Um deles, conforme Nippold (2004: 2), é o de que o desenvolvimento linguístico se revela melhor em textos completos. Para esta autora, os textos narrativos, expositivos ou argumentativos são os mais adequados para a manifestação das marcas do desenvolvimento linguístico, já que os falantes/escritores, nestes textos, entram em ações comunicativas plenas, o que lhes exige um esforço linguístico (e cognitivo) bastante mais substancial do que em estímulos constituídos por pares de perguntas e respostas, por exemplo.

Se é no discurso – enquanto ação comunicativa – que a língua se desenvolve (Hickmann 2003, cf. Tolchinsky 2004: 235), então, para além de completos, os textos devem

também ser genuínos (ou o mais próximo possível, considerando-se o caráter de simulação típico do ambiente escolar). Portanto, o foco na obtenção de textos completos e genuínos tem como uma das suas consequências a possibilidade de fazer emergir das escolhas dos falantes/escritores, nestas situações naturais de comunicação, comportamentos linguísticos específicos, revelando-se, no final, os seus padrões de uso.

A linguística de *corpus* é, deste modo, o caminho a ser seguido, não como um modelo teórico, mas como um modelo metodológico, como um conjunto de instrumentos de análise que permitem ao analista discutir questões que não são transparentes à intuição (Bloor e Bloor 2004: 247). O *corpus* é a fonte natural de extração de dados para a exploração de teorias ou hipóteses, para se poder validá-las, refutá-las ou refiná-las (McEnery e Hardie 2012: 6).

Contudo, adotar um *corpus* também implica afiliar-se num conjunto de convicções teóricas relativas à questão estudada. A mais enfática destas convicções é a de que um estudo baseado num *corpus* lança luz no desempenho linguístico, e não na competência (Leech 1992: 107). Isto significa que o analista volta a sua atenção para a manifestação comportamental da língua em ocorrências naturais na forma de discursos falados, escritos ou multimodais. A partir de um *corpus* representativo, que se constitui num fundamento sólido para a descrição dos usos da língua, tem-se, portanto, a possibilidade de tornar explícitos padrões que emergem do uso. O recurso a métodos que, graças à evolução tecnológica ocorrida nas últimas décadas, permitem associar tanto abordagens qualitativas como quantitativas no tratamento de um grande número de textos garante alguma representatividade dos itens analisados em função das ocorrências naturais na língua, o que se corrobora por meio da aplicação de testes estatísticos, pois que podem assegurar que os resultados encontrados não ocorrem por coincidência. Em particular, é bastante adequada a aplicação de testes de correlação, tipo de análise estatística que consiste em investigar a relação entre dois grupos de variáveis a fim de identificar funções lineares de um grupo que se correlacionam maximamente com as funções lineares do outro grupo (Everitt e Skondrall 2010: 68). Estes testes estatísticos, portanto, alicerçam, com um esperado rigor científico, o caminho

para se alcançar o objetivo nuclear desta dissertação, que é realizar um estudo de correlação entre um conjunto de indicadores de complexidade textual e a progressão escolar.⁵

Ainda como consequência da convicção de que um estudo de *corpus* deve dar atenção ao desempenho linguístico, os estudos de correlação são secundarizados por breves apresentações (baseadas em frequência percentual ou normalizada) de estudos de caso⁶ do uso da língua que se mostraram relevantes nos textos narrativos e argumentativos ao longo do processo de investigação, relevância esta justificada quer por uma ocorrência numericamente expressiva, como acontece com as orações relativas, quer pelo papel recorrente na literatura sobre o desenvolvimento linguístico em idade escolar, como se dá com os pronomes anafóricos.

Tendo em conta tais considerações, constituiu-se, para os estudos de correlação, um *corpus quasi*-longitudinal de textos de registo narrativo (n=122) e argumentativo (n=122), produzidos por alunos iniciantes no 5.º ano (do 1.º ciclo), no 7.º ano (do 2.º ciclo) e no 10.º ano (do 3.º ciclo), em contexto regular de sala de aula, mais especificamente em aulas de português, de quatro escolas da Grande Lisboa, no período de setembro de 2011 a março de 2012. A amostra tem por objetivo representar a progressão nos anos escolares.

⁵ Por serem os indicadores de complexidade textual divididos em dois conjuntos, o da dimensão lexical e o da dimensão sintática, executam-se, por uma questão unicamente operacional, dois estudos de correlação separadamente (cf. capítulos 4 e 5), razão por que, nesta dissertação, eventualmente se fala em “estudos de correlação”, no plural, sem prejuízo do seu uso também no singular, dotado do sentido global.

⁶ O emprego da expressão “estudos de caso” segue a aceção apresentada por Ponte (2006: 106): “É uma investigação (...) que se debruça deliberadamente sobre uma situação específica que se supõe ser única ou especial, pelo menos em certos aspectos, procurando descobrir a que há nela de mais essencial e característico e, desse modo, contribuir para a compreensão global de um certo fenómeno de interesse.”

A partir da extração de dados do *corpus* por meio de tratamento informático manual e automático e da consequente aplicação de testes estatísticos, pretende-se responder a duas questões globais:

a) Qual a natureza das correlações (positiva, negativa ou inexistente) entre os indicadores de complexidade textual e a progressão escolar nos registos narrativo e argumentativo?

b) A força das correlações observadas entre os indicadores de complexidade textual e a progressão escolar é diferente quando se consideram separadamente o registo narrativo e o registo argumentativo?

Com base nestas perguntas, em que se considera, por registo textual, o conjunto de variáveis dependentes (os indicadores de complexidade textual) e a variável independente (a progressão escolar), fundam-se as seguintes hipóteses, também globais:⁷

a) Existe correlação positiva entre os indicadores de complexidade textual e a progressão escolar tanto no registo narrativo, como no registo argumentativo.

b) A correlação observada entre os indicadores de complexidade textual no registo narrativo e a progressão escolar é positiva.

c) A correlação observada entre os indicadores de complexidade textual no registo argumentativo e a progressão escolar é positiva.

d) A força da correlação observada entre os indicadores de complexidade textual no registo narrativo é mais expressiva do que a força da correlação observada no registo argumentativo.

Esta dissertação está organizada como se segue: para além deste capítulo introdutório, seguem-se cinco outros capítulos. No capítulo 2 – Definições operacionais (p. 11) –, faz-se

⁷ Em cada estudo de correlação (capítulos 4 e 5), apresentam-se hipóteses de investigação específicas ao nível de complexidade sob análise.

uma revisão crítica de um conjunto de definições que julgo fundamentais à compreensão deste estudo, dispondo-as na seguinte ordem: complexidade, texto, contexto e registo. Esta revisão norteia-se, sempre que possível, por pressupostos teóricos das teorias linguísticas vocacionadas para as características do texto/discurso, já que o objeto de investigação selecionado é o texto escolar, e os usos linguísticos que nele se realizam. No capítulo 3 – Procedimentos metodológicos gerais (p. 30) –, descreve-se a metodologia de trabalho aplicada ao estudo na totalidade, incluindo-se os procedimentos para a seleção dos participantes e das escolas, os procedimentos de coleta e de compilação do *corpus*, o desenho da base de dados e a descrição dos programas para o tratamento computacional a que os dados foram submetidos. De seguida, nos capítulos 4 e 5 – respetivamente, Complexidade lexical (p. 41) e Complexidade sintática (p. 65) –, apresentam-se os indicadores de complexidade textual e os procedimentos e técnicas de medição específicos para cada tipo de complexidade, discriminando-se quais indicadores constituem as variáveis dependentes de investigação. Em cada um destes capítulos, tem lugar a apresentação dos resultados da aplicação dos testes estatísticos de correlação, sempre ilustrados com o auxílio de gráficos e tabelas, e a análise das correlações identificadas, em que se procura responder às questões de investigação anteriormente enunciadas. No capítulo seguinte, Estudos de caso (p. 98), descrevem-se alguns padrões de ocorrência dos usos linguísticos notáveis nos textos dos alunos. Por fim, no capítulo 7 – Considerações finais (p. 144) – para além de se avaliar sinteticamente este estudo, apontam-se, à luz das vantagens e limitações identificadas, potenciais implicações pedagógicas, desdobramentos pretendidos desta investigação e sugestões de novas pesquisas sobre o desenvolvimento linguístico em idade escolar.

2. Definições operacionais

Como já referido, o objetivo deste estudo sustenta-se numa relação triangular entre indicadores de complexidade textual, registo e progressão escolar. No presente capítulo, faço uma apreciação crítica de um conjunto de definições que este objetivo compreende explícita ou implicitamente. Trata-se das definições de complexidade, texto, contexto e registo. Estas definições, embora comuns na literatura, não são exatamente consensuais. Discuto-as, sempre que possível, dentro dos pressupostos teóricos da linguística textual, porquanto tais correntes têm como propósito maior entender as regularidades linguísticas de textos (Bussmann 2006: 1190), estando, por isto, em linha com os métodos de investigação baseados em *corpus*. Todavia, sempre que necessário, recorre-se ao contributo de outras correntes teóricas, em particular no que respeita às definições de complexidade, que, em virtude da dimensão dos significados que comportam, não se bastam dentro dos limites da linguística textual.

2.1. Complexidade

Entender o fenómeno da complexidade constitui parte da agenda de variados domínios científicos. Estende-se, por exemplo, da preocupação filosófica em explicar, através do princípio da complexidade dos saberes, a mudança de paradigmas das organizações sociais (Morin 2002) até à tentativa de revisão do darwinismo para explicar os saltos evolutivos, que são os processos atípicos e não-graduais que transformam uma simples célula em organismos multicelulares, altamente complexos, ou mesmo em comunidades inteiras (Woese 2004). Na linguística, o tema da complexidade também é recorrente, mas o modo de defini-la é diverso, resultado de diferentes enquadramentos teóricos. A despeito desta diversidade, é possível traçar, a partir de Dahl (2004, 2007, 2009), uma linha divisória que permite pôr, de um lado, a complexidade como um fenómeno subjetivo e, de outro, a complexidade como um fenómeno objetivo.

A complexidade como um fenómeno subjetivo relaciona-se globalmente com a dificuldade de se dominar uma língua. Neste sentido, a complexidade é sinónimo de dificuldade (Fischer *et al.* 2012: 1). Kusters (2003: 6) assim o entende e reitera o carácter subjetivo ao concluir o seguinte:

“... there is no way to define complexity without being specific about to whom a language is or is not complex. In other words complexity is not a simple predicate attributable to language but a relation between two entities: a language and someone who evaluates the language.”

Kusters (2003) assume a ideia de que a complexidade pode ser tratada como um fenómeno relativo, dependente de um sujeito e implicada nos processos cognitivos a que um indivíduo recorre para o processamento de informações. Esta perspetiva ecoa o uso quotidiano do termo, que, além de dificuldade, também pode significar complicação ou obscuridade, uma releitura do facto de que as componentes de um sistema complexo, como a língua, são usualmente numerosas, diversas e dinâmicas, como advogam Larsen-Freeman e Cameron (2008: 1-2). A complexidade subjetiva, *agent-related* nos termos de Dahl (2009: 50), restringe-se aos esforços mentais que uma pessoa ativa no processo de aprendizagem de uma língua.

Em termos mais específicos, a complexidade subjetiva está relacionada com os graus de dificuldade sentida numa dada tarefa, em particular na compreensão textual. Deste modo, a complexidade configura-se nuclear nos estudos sobre a inteligibilidade (*readability*)⁸, cujo objetivo é avaliar o grau de complexidade dos materiais escritos ao serem lidos ou interpretados por aprendizes (Richards e Schmidt 2002: 442). É esta a lógica que fundamenta o programa educacional norte-americano “Common Core State Standards”, ou, mais comumente, CCSS (EUA 2010), que estabelece os conhecimentos e as capacidades de leitura que os estudantes devem desenvolver ao longo dos doze anos de escolaridade obrigatória para que, no fim deste período, sejam capazes de obter sucesso em exames de acesso ao ensino superior e no mercado de trabalho.

Uma das diretrizes dos CCSS consiste em determinar que textos didáticos e paradidáticos são mais adequados a cada ano escolar, graduando-os numa escala de dificuldade (Shanahan 2012). Esta avaliação centra-se num conjunto de indicadores de complexidade

⁸ Embora não seja parte das preocupações deste estudo, vale referir que o termo *readability* parece não ter encontrado lugar assente no léxico português, pelo que se pode encontrar leturabilidade (Cavaco e Várzea 2010), legibilidade (Cavique 2008) ou inteligibilidade (Scarton e Aluísio 2010). Por se tratar de um processo que está relacionado à facilidade/dificuldade de compreensão (Fischer *et al.* 2012: 3), logo um processo de cariz cognitivo, opto pela última tradução.

(“The standards’ approach to text complexity”) divididos em três dimensões: quantitativa, qualitativa e leitor/tarefa (EUA 2010). A dimensão qualitativa da complexidade de um texto refere-se, por exemplo, ao seu significado, ao seu propósito, à sua convencionalidade, à clareza da língua e aos conhecimentos prévios necessários para a sua compreensão. A dimensão quantitativa considera indicadores como a extensão de palavras e frases e a coesão textual. Por fim, a dimensão leitor/tarefa considera, de um lado, fatores como a motivação, o conhecimento e a experiência do leitor em relação ao texto e, de outro, fatores como a finalidade e a complexidade da tarefa atribuída.

Esta perspectiva tripartida proposta nos CCSS revela uma tentativa de interrelacionar a complexidade subjetiva à objetiva, pois admite que as dimensões quantitativa e qualitativa são inerentes aos textos, enquanto a dimensão leitor/tarefa se centra no indivíduo aprendiz (EUA 2010). Porém, a leitura final é sempre a de que há textos que são mais complexos para uns leitores do que para outros (Fisher *et al.* 2012: 3), conformando, afinal, a noção de complexidade à de dificuldade.

A meu ver, interpretar a complexidade como um fenómeno associado à ideia de dificuldade pode trazer nebulosidade, nomeadamente quando a população em causa são crianças e jovens, e pode sustentar tomadas de posição cujo pendor ideológico se mostra detrativo, originado na ideia de que um uso menos complexo da língua pode insinuar algum estágio de inferioridade cognitiva ou sociocultural de um determinado grupo de falantes/escritores. Portanto, a noção de complexidade aqui adotada não corresponde à noção de dificuldade de processamento cognitivo.

Dito isto, interpreta-se a complexidade, aqui, como um fenómeno objetivo, como uma propriedade do sistema linguístico e, conseqüentemente, das suas estruturas, e não algo que lhe é exterior. O princípio geral por trás desta conceção é o de que a complexidade de um objeto está relacionada com a quantidade de informação necessária para o recriar ou especificar, ou, ainda, para medir a extensão da sua mínima descrição possível (Dahl 2007: 38). Desta forma, pode falar-se tanto em complexidade objetiva sistémica, quanto em complexidade objetiva estrutural.

A complexidade objetiva sistémica consiste nas propriedades de uma dada língua passíveis de serem instanciadas pelos seus usuários. Dito de modo ilustrativo: no português, a adjetivação está relacionada com a flexão em género e número (*garoto pequeno/garotos*

pequenos/garota pequena), enquanto, no inglês, tal não acontece (*small boy/small boys/small girl*). Este tipo complexidade objetiva, logo, está na língua, enquanto conjunto de regras que os falantes/escritores de uma língua têm de dominar para produzir e compreender um número indefinido de frases, permitindo reconhecer também incorreções e ambiguidades. Em síntese, a complexidade está nas gramáticas das línguas. Por esta razão, a complexidade objetiva sistêmica é o enquadramento preferido em muitos estudos de comparação entre línguas.

A complexidade objetiva estrutural, que é o suporte particular do meu estudo, também é um fenómeno intrínseco ao objeto, mas, ao contrário da complexidade sistêmica, relaciona-se com as escolhas que um falante/escritor realiza a partir das dimensões estruturais da língua associadas aos contextos de uso, ou, dito de outra maneira, com o conjunto específico de enunciados produzidos pelos falantes/escritores tal como se deteta em *corpora* (Crystal 2008: 357), o que é passível de caracterização por médias de produção linguística no texto de um falante/escritor. Fundamentados nesta noção restrita de complexidade, são muitos os estudos sobre o desenvolvimento linguístico que se apropriam de médias de produção linguística para avaliar o desenvolvimento linguístico de crianças e adolescentes, de que se pode citar Hunt (1965, 1970), Crowhurst (1979, 1980, 1987), Nippold (2004), Berman (2004, 2007), entre tantos outros.

Edmonds (1999), com um foco não restrito à linguística, refere a existência de um conjunto de categorias conceptuais relacionadas com a complexidade e que podem contribuir para a sua descrição. Destas categorias, destaco a variação⁹, que, adaptada ao domínio linguístico, está envolvida nas diferentes formas de se dispor do sistema linguístico em diferentes ocasiões, manifestando-se nas diferentes escolhas que os falantes/escritores instanciam (Biber e Conrad 2009: 4). E o local próprio da variação instanciada é o texto.

2.2. Texto

⁹ Além da variação, Edmonds (1999: 61) cita a extensão, a ignorância, a ordem, o tempo de processamento, a expressividade e a irredutibilidade, entre outras categorias.

As correntes linguísticas de base textual, surgidas a partir da tomada de consciência de que a língua não se resume a frases isoladas, abstratas ou inventadas, têm proposto diversas definições de texto ao longo das últimas quatro décadas. Tais tentativas incluem desde definições ainda arraigadas ao formalismo frásico, como a de Isenberg (1970, cf. Beaugrande 2011: 288), para quem um texto é uma sequência coerente de frases, ou a de Harweg (1968, cf. Beaugrande 2011: 288), em que um texto é uma série de unidades linguísticas constituída por um encadeamento ininterrupto de pronomes, até às definições orientadas por uma perspetiva sócio-discursiva da língua, como a de Brown e Yule (1983: 26), para quem o texto é o registo verbal de um evento comunicativo, ou ainda a de van Dijk (1976: 3), que define texto como uma construção teórica subjacente ao discurso. Ao considerarem-se textos de aprendizes, estas proposições, embora todas meritórias, não se aplicam satisfatoriamente.

Por um lado, as primeiras limitam os analistas à mera transposição dos esquemas analíticos da frase para o texto, unidade maior vista como uma espécie de superfrase, o que não lhes permite examinar certos matizes pragmáticos próprios das aulas de língua, como, por exemplo, o impacto das tarefas de escrita ou o seu carácter de simulação, que podem interferir na produção dos alunos. De facto, os professores, através das tarefas escolares, que são tipicamente eventos de linguagem elaborados para contextualizar e avaliar outro evento de linguagem (*language-within-language*, como sintetiza Sinclair (2004: 66)), exigem dos seus alunos, com frequência, que simulem realidades variadas, às vezes nunca vivenciadas por eles. A título de exemplo, basta referir que as tarefas de escrita das provas de aferição aplicadas pelo Ministério da Educação nos anos de 2003, 2004, 2006, 2008, 2009 e 2012¹⁰ têm a forma verbal *imagine* nos enunciados conducentes à produção de textos por parte dos alunos.

Esta questão é relevante, na medida em que, como mostram Dixon e Stratta (1980, cf. Ruth e Murphy 1988: 257) num estudo sobre o desenvolvimento da escrita escolar de jovens a partir dos 15 anos, escrever uma narrativa imaginária exige muito mais esforços da parte do escritor do que escrever uma narrativa baseada em experiências pessoais, afetando as escolhas retóricas e linguísticas que o escritor faz, razão por que não se pode tomar um texto

¹⁰ <http://bi.gave.min-edu.pt/exames/exames/provas/>

como apenas um conjunto de frases interligadas, sem considerar as circunstâncias da produção textual.

Por outro lado, as definições de Brown e Yule (1983) e de van Dijk (1976) são exemplos de abordagens teóricas que se centram em demasia na dimensão sócio-ideológica da língua, o que pode relegar para um segundo plano o facto de que uma das funções das produções escolares é servir, aos alunos, como instrumentos de aplicação dos recursos gramaticais aprendidos nas aulas e, aos professores, como instrumentos de avaliação da aprendizagem desses mesmos recursos. As Metas Curriculares de Português estabelecidas pelo Ministério da Educação confirmam este facto: “O ensino dos conteúdos gramaticais assenta na instrução formal e deve ser desencadeado a partir de situações de produção oral e escrita ou de textos orais e escritos” (Portugal 2012: 13). Neste sentido, os conteúdos gramaticais, ou o conhecimento explícito da língua, devem ser considerados transversais às competências de leitura e escrita, já que permitem “o controlo das regras e a selecção dos procedimentos mais adequados à compreensão e à expressão, em cada situação comunicativa.” (Portugal 2009: 15-6).

O confronto entre as definições de texto até aqui mobilizadas – de um lado, as que se sustentam nas formas que a língua assume enquanto propriedades do processo de textualização (Isenberg e Harweg) e, de outro, as que se sustentam na dialética entre o uso da língua e os contextos sociais (van Dijk e Brown e Yule) – ilustram o debate contínuo dentro das linguísticas textuais de tentar diferenciar o texto do discurso¹¹. Nesse debate, como sintetizam Baker e Ellege (2011: 30-31), há as propostas que referem o texto como objeto escrito e o discurso como objeto falado, ou ainda o texto como monológico e o discurso como dialógico.¹² Porém, a distinção mais habitual, como observa Christiansen (2011: 34), tende a assumir o texto como a manifestação física do discurso. Widdowson (2004: 8) define o

¹¹ Para Longacre (1983: 337), com um tom que oscila entre o cómico e o crítico, a distinção entre texto e discurso é uma questão puramente geográfica: o discurso na América é o texto na Europa.

¹² Num contexto mais recente, marcado em particular pelo pensamento pós-estruturalista, o termo “texto” é frequentemente usado para referir-se a qualquer objeto de estudo, como, por exemplo, um filme, uma peça publicitária, uma ópera. Esse uso tornou-se popular nomeadamente pelas mãos de Roland Barthes e Jacques Derrida, que defendem a pluralidade de interpretações possíveis de qualquer entidade cultural, verbal ou não.

discurso como uma realidade extralinguística, cuja interpretação depende do contexto em que foi produzida e a que se tem acesso através do texto, a realidade linguística. Wodak (2011: 48) corrobora-o, afirmando ser o discurso um conjunto de padrões e regularidades de conhecimentos, enquanto o texto é uma realização única e específica do discurso.

Hyland e Paltridge (2011: 1) afirmam que o termo *discurso* se refere a um dos mais expressivos conceitos do pensamento moderno, visto que está implicado nos modos como a língua opera o engajamento das pessoas com o mundo e como cria e modifica formações culturais, políticas e sociais. Por mais que tal concepção, alicerçada na mudança de paradigma ocorrida nos estudos linguísticos a partir da década de 60, seja incontestável, também o é a de que a língua se relaciona a praticamente tudo o que existe, razão por que o emprego do termo *discurso* se torna muitas vezes saturado, cobrindo uma vasta gama de acepções, eventualmente antagônicas.

Por isto, e por os objetivos propostos nesta tese não contemplarem questões de natureza social ou ideológica, mantenho-me, sempre que possível, afastado deste debate. Contudo, uma das questões de investigação que levanto relaciona-se com o controle que o contexto de produção pode exercer sobre a composição textual dos alunos. Assim, é incontornável adotar um procedimento de trabalho que permita incorporar fatores do contexto como parte do processo de investigação. Noutras palavras, para uma análise do texto escolar como a que aqui se apresenta, é necessária uma definição de texto que, a um só tempo, permita descrever as escolhas linguísticas codificadas pelos alunos e não descure as interferências do contexto de produção nessas escolhas. Uma proposta para tornar analiticamente exequível a relação entre a realização linguística e o seu contexto encontra-se em Halliday e Hasan (1989: 10), como se evoca na seguinte citação:

The text is a product in the sense that it is an output, something that can be recorded and studied, having a certain construction that can be represented in systematic terms. It is a process in the sense of a continuous process of semantic choice, a movement through the network of meaning potential, with each set of choices constituting the environment for a further set.

Por um lado, tomar o texto como um produto permite um empreendimento de base descritiva sobre a língua codificada; por outro, tomá-lo como um processo permite aceder a informações essenciais (ou à rede de significados potenciais) para interpretar a sua realização. A noção de processo como apresentada por Halliday e Hasan pode, num primeiro

olhar, ser semelhante à de discurso, porém não obriga o analista a comprometer-se com questões sociais ou psicológicas que o termo “discurso” hoje impõe, não ultrapassando, portanto, os limites desta tese, onde, sempre que necessário, se usa o termo “discurso” sem o valor ideológico que lhe é habitualmente atribuído nas linguísticas textuais.¹³

Uma definição bastante recorrente em muitos estudos sobre o texto e que igualmente permite entender o texto como processo e como produto é a de Beaugrande e Dressler (1981: 11), que o definem como uma ocorrência comunicativa que reúne sete critérios de textualidade: coesão, coerência, intencionalidade, aceitabilidade, informatividade, situacionalidade e intertextualidade, sendo os três primeiros centrados no produtor do texto e no próprio texto, e os restantes centrados no recetor do texto. Estes sete critérios são constitutivos dos modos como a comunicação textual opera. Se qualquer um dos critérios, segundo os autores, não é considerado satisfatório, perde-se o carácter comunicativo dos textos, e um texto não-comunicativo será um não-texto.

Sendo a comunicação realizada apenas com a presença concomitante destes sete critérios de textualidade, é justo indagar se os textos escolares podem ser cobertos por essa definição. Schleppegrell (2002: 139) observa que, devido à influência da figura do professor/corretor na escrita do aluno, o aluno commumente opta por valorizar a gramaticalidade em detrimento da informatividade. Logo, de acordo com a definição de Beaugrande e Dressler, muitos dos textos produzidos por aprendizes não devem ser considerados textos, pois falta-lhes informatividade. É razoável questionar também o critério da intencionalidade, por a produção textual na escola ser, em muitos casos, francamente compulsória.

Para fins operacionais, o texto aqui é entendido como uma atualização situada do sistema linguístico. Ou seja, o texto representa as escolhas linguísticas feitas a partir de um repertório potencial, as quais se adequam a um dado contexto. O texto, como metaforizam Neubert e Shreve (1992: 40), é uma ferramenta e, ao mesmo tempo, revela o usuário da

¹³ Há estudos de pendor psicolinguístico que também diferenciam texto como produto e texto como processo, mas, ao contrário do que se vê nas conceções aqui referidas, o processo implica um conjunto de informações de natureza cognitiva que, em especial, dizem respeito ao planeamento do texto.

ferramenta. A adoção desta perspectiva justifica-se por duas razões: por uma razão metodológica, que permite uma conduta proativa em relação à constituição do *corpus* desta investigação, impedindo a exclusão ou a discriminação de textos que, a partir de uma visão restritiva, podem ser considerados como não-textos; e por uma razão analítica (e também metodológica), que permite a correlação entre as escolhas linguísticas realizadas (em oposição ao que é apenas potencial) e os ciclos escolares, revelando, portanto, possíveis indicadores do desenvolvimento linguístico escolar. Trata-se de uma noção, acima de tudo, conciliadora. Veja-se o que diz Beaugrande (2011: 290) em revisão sobre o duo *potencial e atual*:

Today, the ‘text’ is widely defined as an empirical communicative event given through human communication rather than specified by a formal theory. Each such event ‘rides on’ a dynamic dialectic between the ‘virtual system’ of language (the repertory of possibilities) and the ‘actual system’ constituted by the choices of the text producer; the text is thus on neither side of language versus use, but integrates and reconciles the two.

O texto é o espaço próprio para a identificação das formas linguísticas selecionadas pelos falantes/escritores de uma língua e, no caso dos alunos-escritores, esta seleção tem origem num repertório linguístico, que está sob sistemática reelaboração através do processo de escolarização a que os alunos estão submetidos, tornando-se, neste sentido, progressivamente mais complexos. Para Halliday (2007: 274), durante a vida escolar, um aluno constrói um manancial de recursos que lhe permite, sempre que necessário, criar significados, que se armazenam na sua mente em forma de potencial, porque podem ou não ser ativados e porque, se ativados, podem ou não sê-lo do mesmo modo como foram inicialmente armazenados. O armazenamento da informação estabelece-se, portanto, em forma de sistema potencial, que se define do seguinte modo (Halliday 2007: 274):

It is a language, or some specific aspect of language, like the language of science, in the form of a potential, a resource that you draw on in reading and writing and speaking and listening – and a resource that you use for learning with.

O sistema potencial torna-se tangível, ou, nos termos de Halliday e Matthiessen (2004: 26), é instanciado na forma de texto. Dito em forma de exemplo: quando um professor de Português, em sala de aula, atribui aos seus alunos uma tarefa de produção textual (contar como foram as férias escolares, por exemplo), obtém como resultado tantas instanciações

quantos alunos há em sala, ou seja, um texto é uma particularização das possibilidades de realizações a partir do sistema linguístico. Cada aluno faz a sua própria seleção, instancia o sistema ao seu modo, consciente ou inconscientemente, e esse modo, em princípio, é expressivamente diferente dos modos de seleção feitos pelos seus colegas, embora todos tenham usado o mesmo sistema como fonte de recursos para a instanciação. Um dos fatores que tornam os textos dos alunos semelhantes entre si é a tarefa proposta pelo professor, que lhes faculta um conjunto de informações – o assunto a ser tratado, os interlocutores envolvidos no evento textual e a feição do texto a empregar. Tais informações configuram o contexto, que, conforme Cloran (1999: 177), sempre interfere na codificação da língua, razão por que é fundamental definir o que é contexto, examinando ainda como se relaciona com a materialidade do texto.

2.3. Contexto (e cotexto)

Nas palavras de Widdowson (2004: 8), um texto sem contexto é inerte, mas uma definição operacional de contexto tem por obstáculo a imprecisão com que este conceito tem sido tratado, o que se ilustra facilmente com algumas definições recorrentes na literatura, como a de Crystal (2008: 109), segundo o qual o contexto são as características do mundo não-linguístico em relação às quais as características linguísticas são sistematicamente usadas, ou a de Bussmann (2006: 245), para quem o contexto se refere a todos os elementos, verbais e não-verbais, de uma situação comunicativa, ou ainda a de Halliday e Webster (2009: 240), que dizem ser o contexto o ambiente extralinguístico do texto. Definições desta natureza corroboram o ceticismo de Widdowson (2004: 36), que conclui que o contexto é bastante esquivo a definições.

Há, porém, esforços no sentido de não apenas defini-lo, mas também de inscrevê-lo nas linguísticas textuais como uma categoria descritivo-analítica, afastando-se, inclusive, a ideia de que o contexto é um “saco azul”, onde tudo se mete, sem a necessidade de mais explicações. Tais esforços, empreendidos, por exemplo, por Auer (1996), Hasan (1996), Linell (1998), Fetzer (2004), Halliday (2007), van Dijk (2007, 2008) e Widdowson (1973, 2004), oferecem tratamentos da noção de contexto que, mesmo enquadrados por pressupostos teóricos distintos, apresentam muitas semelhanças, possibilitando a um analista não apenas extrair uma definição de trabalho que seja integrada e coerente com a definição de texto (e registo, como se verá na secção seguinte) que pretende adotar, mas também que seja

ajustada às particularidades da produção textual em foco, que, nesta tese, é a escolar. Para tanto, segue-se uma exposição sintética desses tratamentos, que representam, *grosso modo*, o contexto como um constructo teórico bidimensional, distintos em cotexto e contexto.¹⁴

O cotexto, como batizado por Catford (1965, cf. Bussmann 2006: 261), é um recurso bastante invocado em estudos das linguísticas textuais, principalmente em estudos baseados em *corpora*. Por isto mesmo, muitas vezes, é tratado como parte do conhecimento comum da linguística, portanto sem a necessidade de explicá-lo. Van Dijk (2007: 3) assegura que o cotexto, ou contexto verbal, como prefere, não carece de tratamento especial, já que um dos principais objetivos da análise do discurso (variação da LT) é justamente estudar os ambientes discursivos de palavras e frases. No entanto, um rápido exame das poucas definições disponíveis revela o contrário:

... both prior language in a communication (the writing or speech that comes before the utterance in focus) and what follows it. (Hewings e Hewings 2005: 19)

... the set of other words used in the same phrase or sentence, also called the linguistic context” (Yule 2010: 285)

... the words, phrases, clauses or sentences that come before and after [a piece of text]. (Jones e Lock 2011: 138)

Nas definições acima, não há acordo quanto ao raio de ação do cotexto (o sintagma ou a frase, para Yule; ilimitado, para Hewings e Hewings), quanto à sua constituição (um conjunto de palavras, para Yule; qualquer unidade gramatical, para Jones e Lock) ou mesmo quanto ao que está circundado pelo cotexto (um enunciado, para Hewings e Hewings; uma

¹⁴ Autores há, em particular os de tendência sócio-semiótica, como Halliday (2007) e van Dijk (2008), já referidos, que distinguem uma terceira dimensão contextual, a cultura, que consiste na dimensão de códigos, práticas e padrões partilhados por uma comunidade, fornecendo o *background* necessário ao enquadramento social. O conceito de cultura, neste caso, não se restringe ao conceito popular, que habitualmente se refere à cultura como algo relacionado com as origens étnicas ou as ideias e os hábitos do passado, mas com tudo aquilo que possa ser relevante enquanto procedimento social apropriado a uma atividade linguística. Por o conceito de cultura não estar relacionado com o foco desta investigação, não o pormenorizo.

porção de texto, para Jones e Lock). Mais incerta torna-se a discussão sobre o cotexto se se considerar outra afirmação de van Dijk (2008: 117): “I prefer to study ‘cotext’ as ‘previous parts’ of dynamically defined text or talk itself...”, em que a direção da abrangência do cotexto se limita unicamente ao que já foi dito ou escrito, não incluindo o que está por vir. Fetzer (2004: 5) garante que tanto a realização linguística prévia ao fenómeno estudado, como a realização posterior a ele constituem o cotexto. Isto exemplifica-se nas relações coesivas, em que um determinado item linguístico tem o seu significado plenamente satisfeito pela recuperação de outro item linguístico, que tanto pode estar localizado em posição anterior (recuperação retrospectiva), como em posição posterior (recuperação prospetiva). Também, como defende Auer (1996), nas atividades conversacionais, toma-se por indiscutível que um turno conversacional, além de estar ancorado no anterior, também configura o seguinte, constringendo-lhe a forma e a função. Nas atividades conversacionais, a enunciação de uma frase depende do contexto linguístico existente para a sua produção e interpretação e torna-se, por sua vez, um evento que modela uma nova frase, ou seja, um novo contexto linguístico que se seguirá (Fetzer 2004: 6).

Um estudo que tem no cotexto um dos seus fatores de análise das ocorrências linguísticas necessita de uma definição operacional integrada, razão por que trato o cotexto como o conjunto de ocorrências linguísticas relevantes para outra ocorrência linguística, em que esta mantém com aquele conjunto, na mesma superfície textual, alguma relação, de natureza retrospectiva ou prospetiva. Por outras palavras, o cotexto é a dimensão variável, portanto dependente do fenómeno a ser investigado, em que uma dada ocorrência linguística contribui para a existência de outra, ambas localizadas no mesmo texto. Entendida desta forma, a noção de cotexto sujeita-se, sem restrições, à aplicação das várias técnicas de análise baseadas em *corpora*. Não quer dizer isto que o cotexto se transforme unicamente numa periferia linguística, como pode ser interpretado, muitas vezes, em estudos descritivos de *corpora*, em que há commumente referência ao cotexto como a gama de palavras à esquerda e à direita de uma outra palavra. Como lembra Linell (1998: 144), no cotexto, ocorrem processos de recontextualização, que implicam numa transferência/transformação de informações dentro do mesmo texto. Também Widdowson (1973: 57) enfatiza a importância do cotexto, ao tomá-lo como categoria analítica para proceder ao que chama de abordagem intratextual, que serve para identificar as propriedades formais que legitimam a interrelação de frases que as harmoniza como um texto.

Inspecionar o cotexto implica considerar o texto na sua materialidade, portanto como um produto, como o espaço próprio para a identificação de padrões de realização linguística, transformando-se numa categoria imprescindível para os estudos baseados em *corpora*, os quais, através do processamento de grandes quantidades de textos, apreendem as regularidades dos usos linguísticos para atribuí-las como particulares a determinadas variedades (ou registos). Porém, o foco no cotexto, pela abordagem intratextual, é restritivo, já que tende a excluir os fatores de natureza psicológica, social ou pragmática que constituem a conformação discursiva do texto, o que pode resultar em análises parciais do material linguístico em estudo, como acontece, por exemplo, nas relações coesivas, que, em muitos casos, dependem de uma interpretação de cariz pragmático para a sua inequívoca resolução (Cornish 2009: 97). Para solucionar esta questão, a abordagem intratextual deve complementar-se com uma abordagem extratextual, isto é, com a inclusão do contexto de situação.

Diferentemente do que acontece com o cotexto, o contexto de situação, ou simplesmente contexto, tem sido tópico de um intenso debate ao longo das últimas décadas, como se vê, particularmente, em Halliday e Hasan (1989), Hasan (1996, 1999), Widdowson (2004), Fetzer (2004), Halliday (2007), van Dijk (2008). O debate justifica-se na convicção de que o contexto é o *locus* do texto (Halliday e Hasan 1989: 12), sendo, portanto, consensual entendê-lo como o entorno relevante numa interação linguística, como sistematizado por Hasan (1999: 232), que se ampara na tríade Bronislaw Malinowski, John Rupert Firth e Michael Halliday para defender que nenhum uso real da língua está separado da situação em que foi produzido.

O contexto, portanto, consiste no conjunto de particularidades – objetos, pessoas, eventos, etc. – relevantes para uma interação, tornando-se uma força dinâmica tanto para a criação como para a interpretação de textos (Hasan 1996: 41). Trata-se de uma força que permite predizer quais são os elementos necessários à estrutura textual a ser produzida num dado evento. A possibilidade de predição é explicada pelo controlo que o contexto de situação aplica aos significados produzidos num texto. Eggins (1994: 9) reforça este argumento ao afirmar que o contexto está no texto: o texto traz em si os aspetos do contexto no qual foi produzido e, presumivelmente, onde é considerado adequado. Contudo, tal controlo não ocorre numa única direção. Do mesmo modo, se conhecidos os elementos textuais, é possível a um ouvinte/leitor depreender qual o contexto de situação em que o texto foi produzido. Hasan (*idem*: 42) compendia esta relação na seguinte citação: “... if we have access to the

context, we can predict the essentials of the text; if we have access to the text, then we can infer the context from it.”

Conclui-se daqui que é preciso que a relação entre o texto e o contexto seja tratada como dialógica e, conseqüentemente, que se entenda o modo como estas duas entidades interagem no processo de construção de significados, visto que uma é o instrumento de desenvolvimento da outra. Esta concepção dialética da relação entre texto e contexto reflete pressupostos da Sociologia da Educação, de Bernstein (1999), que assegura que a análise que toma como ponto de partida as propriedades internas da forma do texto revela a interdependência entre essas propriedades e o contexto social no qual elas são ativadas e legitimadas. Neste sentido, quaisquer que sejam as escolhas gramaticais e as escolhas das formas de expressão do significado, conscientemente ou não, são sempre regularizadas pelo meio, numa espécie de jogo de coerência instituído pela comunidade discursiva, através de códigos semióticos, que, na sua essência, servem para representar significados entre os membros dessa comunidade.

Durante um ato comunicativo, a percepção que o falante/escritor tem do contexto permite-lhe ativar (consciente ou inconscientemente) as escolhas de expressão de significado no texto, do mesmo modo que os significados escolhidos (re)constroem o contexto. Esta reciprocidade evidencia o caráter probabilístico do uso da língua. Segundo Hasan (1996: 39), há três variáveis contextuais que caracterizam qualquer processo interativo. São elas: i) a natureza do processo social, ou o que se quer atingir através de atos de interação verbal; ii) a natureza da relação entre os interlocutores, ou quais os papéis sociais que se estabelecem entre esses interlocutores durante a interação; e iii) a natureza do modo de transmissão da mensagem, ou qual o papel da língua na estrutura do evento interativo. O exame destas três variáveis garante um entendimento abrangente de um ato comunicativo, de acordo com Hasan (2009: 172), que propõe os seguintes termos técnicos para se referir às mesmas: i) campo, ii) relações e iii) modo.

É preciso ponderar que, quando se descreve um contexto de situação, não se está a descrevê-lo como constituição física, ou seja, não se pode tomá-lo como uma captação objetiva da realidade. O contexto de situação é uma abstração, um aparato teórico, sem existência em si mesmo, como fica explícito na citação abaixo (Hasan 1995: 191):

Context of situation, as category of linguistic analysis, was analogous to other linguistic categories (...). They are theoretical fictions; they do not name any physical object having existence apart from the theory itself. Much less could they be regarded as phenomena of primary experience.

Quer isto dizer que o contexto de situação, enquanto dimensão extralinguística, reporta-se a elementos abstraídos da realidade imediata, filtrados pela mente dos interlocutores (e do analista linguístico). É esse filtro que dá a medida do que é relevante ou não ao desenvolvimento de um evento linguístico, indicando os caminhos pelos quais a interação verbal pode continuar, o que torna cada contexto de situação único, criado por interlocutores também únicos que se pensam o centro do universo (Hasan 1996: 38).

O contexto de situação, enquanto dimensão extralinguística, reporta-se a elementos abstraídos da realidade imediata, filtrados pela mente dos interlocutores (e do analista linguístico). É esse filtro que dá a medida do que é relevante ou não ao desenvolvimento de um evento linguístico, indicando os caminhos pelos quais a interação verbal pode continuar, o que torna cada contexto de situação único, criado por interlocutores também únicos que se pensam o centro do universo (Hasan 1996: 38). Esta compreensão de contexto aplica-se, com bastante eficiência, a muitos tipos distintos de interações linguísticas, mas, ao considerar-se um texto produzido por um aluno em sala de aula a partir de uma tarefa de escrita, é pertinente, e até mesmo estratégico para fins metodológicos, perguntar qual é o contexto de situação. Dito de modo ilustrativo: se a um aluno é pedido que escreva um texto narrativo em que conte, em primeira pessoa, como seria viver como um objeto de escrita durante 24 horas¹⁵, qual é o contexto de situação? A partir de Halliday (1978: 228), que sugere que a feição de uma análise contextual depende dos interesses de pesquisa, pode-se dizer que há muitas respostas possíveis, das quais conjeturo, sinteticamente, duas. Num primeiro caso, se o que se pretende investigar se relaciona com as práticas pedagógicas nas quais a língua funciona como meio de compreensão de conteúdos, o contexto de situação pode configurar-

¹⁵ Faço referência à tarefa de escrita constante da prova de aferição aplicada aos alunos do 1.º ciclo do ensino básico no ano de 2008: (http://bi.gave.min-edu.pt/exames/download/Prova_de_Afericao_LP_1__ciclo_08.pdf?id=206).

se do seguinte modo: i) campo: atividade de aprendizagem de uma língua; ii) relações: professor e alunos; iii) modo: interação pelo texto escrito como aparelho de instrução para tal aprendizagem.

Noutro caso, se o que se pretende investigar são, por exemplo, os dispositivos de coesão pronominal selecionados pelo aluno a configuração contextual aponta para as atividades, pessoas, situações, enfim tudo o que possa tomar lugar dentro da narrativa imaginada pelo aluno. Assim, num tom quase hipotético (já que cada aluno faz as suas próprias abstrações a partir da tarefa, as quais podem ou não ser semelhantes com as de outros alunos), tem-se: i) campo: vivência de 24 horas como uma caneta; ii) relações: caneta/lápis/borracha/escritor; e iii) modo: oral/escrito através de diálogos. É, portanto, a tarefa o que define a tónica do texto pretendido, pelo que o aluno se orienta por ela, de modo a permanecer conectado com o núcleo de informações que ela propõe/impõe, para assim ser bem-sucedido, mesmo que isso tenha um custo, como, por exemplo, uma excessiva repetição lexical de itens presentes na tarefa (Martins 2008). Apesar de hipotética (mas bastante plausível e até expectável), esta configuração contextual desafia a definição de contexto enquanto ambiente extralinguístico (Halliday e Webster 2009: 240), já que as variáveis contextuais (campo, relações e modo) habitualmente estão, total ou parcialmente, verbalizadas numa tarefa de escrita, razão por que, nesta tese, adoto a tarefa de escrita como o principal contexto de situação dos textos do *corpus*. Isto, no entanto, não implica que os significados num texto escolar não sejam também suportados no contexto de situação extralinguístico, no seu sentido clássico. Logo, quando necessário e para efeitos de distinção, chamo ao primeiro contexto de situação local e, ao segundo, contexto de situação global.

Por terem influência direta e expressiva no tipo de linguagem que se produz num evento linguístico, como garante Eggins (1994: 53), o contexto de situação (neste caso, local) constitui-se em ferramentas de trabalho bastante recorrentes nas linguísticas textuais, porque uma configuração contextual específica em regular associação com uma configuração linguística específica caracterizam uma variedade funcional da língua, ou um registo,¹⁶ tópico a ser considerado na próxima secção.

¹⁶ Como lembram Quirk *et al.* (1985: 16), este tipo de variedade funcional da língua é centrada no uso da língua, diferentemente dos outros tipos de variedades, que são centradas no usuário da língua.

2.4. Registo

Para satisfazer o objetivo desta tese, que é analisar a natureza da correlação entre indicadores de complexidade textual e progressão escolar em dois registos escolares – narrativo e argumentativo –, comparando-se-os entre si, é impositivo não somente definir o que se entende por *registo*, já que é o recorte de uso da língua aqui selecionado, mas também demarcá-lo de outros termos, em particular do género, cuja aceção se confunde com a de registo, sendo muitas vezes empregues como sinónimos, como lembra Lee (2001: 41).

Neste sentido, é exemplar a definição de McEnery e Hardie (2012: 250-1), para quem se podem chamar registos *chats online*, artigos científicos ou novela:

A way of classifying texts according to non-linguistic criteria, such as the purpose for which a text was produced, the intended audience, the level of formality, whether its purpose is narration or description and so on. For most purposes, the term *genre* can be considered equivalent, though some authors differentiate them.

Aqui, por se tratar de um estudo que se sustenta em métodos e técnicas da linguística de *corpus*, a opção natural seria manter-se em linha com a definição corrente nesta área da linguística. No entanto, na definição de McEnery e Hardie, em que os termos *género* e *registo* são intercambiáveis, estão subjacentes pressupostos teóricos de uma linguística de base funcional, dado que se classifica o género a partir de critérios extralinguísticos, que a afastam do tipo de investigação aqui assumido, que é fundamentalmente linguístico. Tais pressupostos respaldam-se na definição que Martin e Rose (2003: 8), representantes da chamada Escola de Sydney,¹⁷ oferecem para o género: “[Genre is] a staged, goal-oriented social processes through which social subjects in a given culture live their lives.” Nestes autores, diferentemente do que se vê naqueles, o género, que se substancia na forma de tipos de textos (exposições, reconto ou anedotas), é um construto teórico diferente de registo, o qual consiste num conjunto de configurações extralinguísticas que permitem a análise dos contextos de situação em que ocorrem os tipos de textos (Martin e Rose 2003: 296).

¹⁷ Para aprofundamentos sobre as as quatro correntes mais populares de estudos de género (Escola de Sydney, Nova Retórica, Inglês para Fins Específicos e Tradição Francófona), consulte-se a detalhada obra de Bawarshi e Reiff (2010).

Entende-se, em concordância com Lee (2001: 38), que a diferença entre género e registo é apenas teoricamente motivada, e, por isto mesmo, *ténue*. São dois modos distintos de se analisar o mesmo objecto textual. Isto quer dizer que, por um lado, pode-se analisar um texto pela perspectiva do registo, o que implica analisá-lo como um objecto linguístico, isto é, como o local de instanciação de uma dada configuração linguística associada a um dado contexto de situação (contexto de situação como descrito na secção anterior, dividido, nesta tese, em contexto global e contexto local). É a variedade de acordo com o uso, pelo que diferentes situações requerem diferentes configurações linguísticas. O exame que se faz do registo de um artigo científico, por exemplo, considera quais são os recursos linguísticos - pronomes, verbos auxiliares, orações dependentes, nominalizações, etc. - que nele são mais frequentes, além de funcionalmente determinados pelo contexto imediato da sua produção.

Por outro lado, analisar o texto pela perspectiva do género implica analisá-lo como um objecto sociocultural, ou seja, como um evento interativo em que estão em jogo os usos ideologicamente emoldurados que se podem fazer desse objecto. As palavras de Martin e Rose (2003: 8), anteriormente citadas, são ilustrativas desta perspetiva. Neste sentido, é o género que permite a participação numa dada comunidade. Em termos teórico-metodológicos, pela perspectiva do género, o mesmo artigo científico será analisado na sua organização retórica, em que se analisam também os seus estágios constitutivos (resumo, introdução, métodos, etc.) - elementos pontuais do género, ou seja, não recorrentes -, e no seu papel social, nas motivações de poder e dominação da parte do escritor e na potencial submissão a que o receptor se expõe na recepção que faz do texto, por exemplo.

O género, enquanto artefacto sociocultural, não é um conceito pertinente para esta tese, que se centra na realização de características linguísticas dos textos escolares que podem ser tomadas como indicadores de desenvolvimento ao longo da progressão escolar, e não em questões de cariz sociocultural ou retórico. Assume-se, ao contrário disto, a análise do texto enquanto registo, o que, reitero, se reflecte nas propriedades linguísticas dum dado texto constrangidas por factores contextuais, além de ser convergente com a definição de texto já apresentada, entendido como uma atualização situada do sistema linguístico.

O termo “registo”, em linha com Biber (2006: 11), refere uma variedade linguística que é contextualmente determinada e cujas realizações lexicais e sintácticas, podem ser tomadas como propriedades que a distinguem de outras variedades. Ou seja, tratar do registo

de um tipo de textos, como, por exemplo, a notícia, o artigo científico ou o conto policial, é tratar dos modos como a língua é empregue nesses textos e, consequentemente, dos modos como ela serve para caracterizá-los, tornando-os adequados aos seus contextos de uso e, ao mesmo tempo, contrapondo-os a outros tipos de textos, a outros registos.

Nesse sentido, a um trabalho de investigação sobre textos escolares, é fundamental a ideia de que, apesar de sermos capazes de adquirir, sem nenhuma instrução explícita, um conjunto amplo de diferentes registos falados, é apenas na escola que nos são ensinados os registos escritos, que se configuram com diferentes padrões linguísticos, a depender dos diferentes contextos de uso, tornando-se mais e mais exigentes ao longo da progressão na escola (e na universidade). Como afirmam Biber e Conrad (2009: 3): “Success requires learning the particular language patterns that are expected for particular situations and communicative purposes”.

Inspirado em Halliday, McIntosh e Stevens (1964), o conceito de registo envolve a variação linguística que é dependente do uso, ou da situação, ou ainda da ocupação profissional¹⁸, como lembram Richards e Schmidt (2002: 452), o que, muitas vezes, acaba por ser o critério primeiro na distinção dos registos, pelo que se pode falar em registo académico, jornalístico ou escolar. Quer isto dizer que os falantes/escritores, quando produzem textos num mesmo registo, estão a realizar tarefas comunicativas semelhantes, de que resultam, como assegura Biber (1995: 1), semelhanças linguísticas, que, afinal, também podem ser tomadas como critérios distintivos doutros registos, obtendo-se uma classificação mais fina de tipos, tais como novelas, cartas de reclamação, editoriais, artigos científicos, etc. É com base nesta perspectiva, a de que os falantes/escritores, quando se ajustam a um registo, estão a ajustar-se às circunstâncias do contexto e aos propósitos comunicativos e, consequentemente, a um padrão lexical e gramatical, que se entende que a narração e a argumentação escolares são registos especializados, ou sub-registos, do registo escolar, e, nesta condição, podem ser comparados.

¹⁸ A variação pelo registo opõe-se à variação pelo dialeto. Enquanto este implica o que uma pessoa fala em virtude do que ela é, aquele implica o que uma pessoa fala em virtude do que ela está fazendo (Webster 2009: 445).

3. Procedimentos metodológicos gerais

Neste capítulo, apresento os passos metodológicos seguidos nesta investigação, aplicáveis genericamente aos estudos de correlação sobre a complexidade textual. Para tanto, faço, inicialmente, uma caracterização global do estudo, em que justifico a opção pelo *corpus quasi-longitudinal*¹⁹. Segue-se uma exposição sobre as escolas e os participantes selecionados para esta investigação, os estímulos de obtenção dos textos, bem como as questões ético-legais que envolveram esta fase da investigação. Por fim, apresento as fases de compilação do *corpus*, nomeadamente a transcrição, o armazenamento e a anotação, com a descrição dos programas informáticos que estão a ser utilizados.

3.1. Caracterização do estudo

De acordo com Reppen *et al.* (2002: vii), uma descrição adequada do uso da língua deve estar alicerçada numa análise de carácter empírico de textos naturais, devendo tal análise realizar-se a partir de grandes quantidades de textos, produzidos por diferentes escritores/falantes, de modo que as potenciais conclusões não sejam influenciadas pelas idiosincrasias de uns poucos participantes. Satisfaz este tipo de descrição, em termos metodológicos, um estudo baseado em *corpus*, em que se emprega um conjunto de procedimentos investigativos cujo fundamento é permitir ao analista, a partir dos dados obtidos em grandes coleções de textos, explorar teorias ou hipóteses a fim de validá-las, refutá-las ou refiná-las (McEnery e Hardie 2012: 6).

Para o presente estudo, que tem por objetivo central explicitar a correlação entre indicadores de complexidade textual e progressão escolar a partir dos usos linguísticos em dois registos escolares (narrativo e argumentativo), os métodos de trabalho da linguística de *corpus* são os instrumentos adequados, para o que se constituiu um *corpus* de textos produzidos por alunos monolingues de português procedentes de cada um dos três ciclos do ensino básico em Portugal. Trata-se, portanto, de textos de alunos com diferentes idades e níveis de escolaridade situados num estágio específico da sua progressão escolar.

¹⁹ O *corpus* deste estudo está disponível online para consulta por linhas de concordância em: <http://alfclul.clul.ul.pt/CQPweb/codes/>

Além de uma abordagem empírica, os trabalhos de investigação sobre o desenvolvimento linguístico em idade escolar sob os efeitos da progressão escolar idealmente estruturaram-se como estudos de carácter longitudinal, em que o mesmo procedimento de coleta dos dados se repete ao longo de uma faixa de tempo e em contextos semelhantes. Porém, este tipo de estudo revela-se muito exigente, devido às dificuldades de se manter a mesma população em constante observação e ao consumo de tempo que estudos desta natureza implicam. Uma solução é simular, num estudo transversal, o aspeto temporal (Litosseliti 2010: 58), o que significa obter, no mesmo tempo e nas mesmas condições, dados que representam diferentes estádios de desenvolvimento do fenómeno que se quer escrutinar, configurando-se, portanto, uma evolução aparente no tempo. Tal procedimento elimina as dificuldades próprias dos estudos longitudinais e permite fazer inferências sobre o desenvolvimento temporal de um fenómeno. Johnson e Johnson (1999: 93) asserem que um estudo transversal que observa um grupo de aprendizes em diferentes níveis de escolaridade exhibe efeitos de estudos longitudinais. Granger (2008: 263) chama a estes estudos de *quasi*-longitudinais, nomenclatura que sigo. O *corpus* “College Learners’ Spoken English Corpus”, um dos mais utilizados como referência em pesquisa sobre o desenvolvimento linguístico de aprendizes conforme Cock (2010: 126), é um bom exemplo de um *corpus quasi*-longitudinal. Quanto aos estudos baseados em *corpora quasi*-longitudinais, podem-se citar, como exemplos, Bos (2001), que analisa a aquisição da temporalidade em produções narrativas de crianças bilingues (árabe e alemão) imigrantes nos Países Baixos, e Woods (2000), que explora as mudanças na produção vocálica no dialeto de três gerações de famílias.

Portanto, a fim de se obter um conjunto de textos que configurassem um *corpus quasi*-longitudinal e que permitissem avaliar as escolhas linguísticas enquanto dependentes dos contextos de situação, em particular da tarefa de escrita (o contexto mais imediato e motivador da produção textual na escola – ambos os tópicos já discutidos no capítulo Definições operacionais), foram aplicadas a todos os participantes, independentemente do ciclo de ensino, as mesmas duas tarefas de escrita em situações semelhantes de sala de aula.

3.2. Escolas e participantes

A fim de avaliar a correlação entre o desenvolvimento linguístico e a progressão nos anos escolares, previamente instituiu-se como referencial de progressão cada um dos três ciclos da escolaridade básica do sistema educacional português. Deste modo, selecionaram-

se alunos regularmente matriculados no 5.º ano, a representar o desenvolvimento linguístico ocorrido no 1.º ciclo; alunos matriculados no 7.º ano, a representar o desenvolvimento linguístico ocorrido no 2.º ciclo; e alunos matriculados no 10.º ano, a representar o desenvolvimento linguístico ocorrido no 3.º ciclo.

O acesso às escolas, a aplicação dos estímulos, a obtenção e o consequente manuseio para fins investigativos dos textos realizou-se sob uma perspectiva ético-legal, pelo que se seguiram os procedimentos para investigação em meio escolar regidos pelos termos do Despacho N.º 15847/2007, publicado no DR 2ª série nº 140, de 23 de julho, e concebido especificamente para regular a aplicação de inquéritos em meio escolar. Fundamentado por este despacho, o acesso às escolas foi aprovado pela Direção Geral de Inovação e Desenvolvimento Curricular (DGIDC), sob o número de processo 0260700001.

Após a efetiva aprovação para o acesso, deu-se início à fase de recrutamento de participantes, para o que foram selecionados treze agrupamentos de escolas de ensino básico e secundário localizados no distrito de Lisboa. A escolha desses agrupamentos foi aleatória. No contacto inicial, por correio electrónico, foi apresentada aos diretores a autorização de acesso às escolas, obtida junto da DGIDC, com os seguintes documentos em anexo: plano de trabalhos, em que se descreviam os estímulos e os procedimentos para a sua aplicação e consequente coleta do *corpus*, carta de recomendação de um dos orientadores da tese e projeto de investigação. Do total de diretores, seis não responderam, três responderam negativamente e quatro responderam positivamente. Das quatro escolas que concordaram em participar do estudo, três estão localizadas no concelho de Lisboa, nomeadamente nas freguesias de São João de Deus, São Sebastião da Pedreira e Santa Maria dos Olivais, e uma localizada no concelho de Torres Vedras, na freguesia de São Pedro e São Tiago. Todas são escolas públicas. Enquanto documento de garantia da boa consecução do plano de trabalhos e da definição de papéis das partes envolvidas neste processo, firmou-se entre a Direção de cada agrupamento e o investigador o Protocolo de Cooperação (Anexo 9.1). Nele, descreve-se um conjunto de medidas para a manutenção da ética necessária a uma investigação de natureza observacional, como a que aqui se propõe, concretamente o anonimato da escola, dos docentes e dos discentes envolvidos.

Tendo sido definidas as escolas, os diretores decidiram autonomamente que classes seriam incluídas no estudo. A partir de então, a direção de cada agrupamento ficou encarregada de enviar o Termo de Consentimento (Anexo 9.2) aos encarregados de educação, a

solicitar-lhes autorização para a participação dos educandos em causa nos trabalhos investigativos. Assim, iniciou-se uma série de reuniões entre o investigador, os diretores das escolas e os professores das turmas escolhidas com o objetivo de se esclarecerem eventuais dúvidas que tivessem restado após a leitura do plano de trabalhos e do projeto de tese e também com o objetivo de reforçar, junto dos professores, o compromisso quanto aos procedimentos da aplicação dos estímulos, em particular quanto à não interferência na produção dos alunos-participantes, nem com sugestões, nem com esclarecimentos de qualquer tipo. Em respeito ao calendário letivo de cada escola, os professores ficaram responsáveis por escolher as datas de aplicação dos estímulos que melhor se adequassem ao andamento das aulas, com a única exigência de que tal aplicação decorresse entre os meses de setembro de 2011 e janeiro de 2012, período relativo aos três primeiros meses do ano letivo de 2011/2012. A escolha desse período não foi aleatória. Pelo facto de esta investigação estar focada em produções textuais que pudessem revelar indicadores de desenvolvimento linguístico, comparando-se os ciclos de ensino, optou-se pela seleção de alunos recém-iniciados nos níveis de ensino imediatamente posteriores ao último ano de cada ciclo sob análise, evitando-se, deste modo, as interferências do novo período.

Nippold (2004: 2), como já referido, afirma que o desenvolvimento linguístico se revela melhor em textos completos, pelo que os textos narrativos, expositivos ou argumentativos são os mais adequados para a manifestação das marcas do desenvolvimento linguístico, tendo em conta que os falantes/escritores, nestes textos, entram em ações comunicativas plenas, o que lhes exige um esforço linguístico (e cognitivo) bastante mais substancial do que em estímulos constituídos por pares de perguntas e respostas, por exemplo. Deste modo, foram aplicados a cada aluno dois estímulos, apresentados abaixo, um para a obtenção de um texto narrativo e outro para a obtenção de um texto argumentativo:

Estímulo 1 (tarefa de escrita de um texto narrativo):

Narra um facto marcante (real ou imaginado) que tu e teu(tua) melhor amigo(a) viveram durante o último verão.

Estímulo 2 (tarefa de escrita de um texto argumentativo):

Achas que as redes sociais (Facebook, Twitter, Google+, Windows Live Space, etc.) são importantes hoje em dia? Escreve um texto para ser publicado no blogue da tua escola em que exponhas a tua opinião sobre as redes sociais. Neste texto deves dizer se és a favor ou contra a existência das redes sociais. Não te esqueças de justificar a tua opinião!

Como estratégia de normalização dos textos, solicitou-se a todos os alunos que escrevessem as redações a caneta azul ou preta, que não atribuísem títulos e que limitassem a extensão do texto a, no mínimo, 20 linhas, e, no máximo, 25.²⁰ Também, para garantir que estas tarefas fossem tomadas pelos alunos como semelhantes às que fazem parte das suas práticas escolares, evitando-se situações interpretadas como atípicas, elas foram submetidas à apreciação independente de três professores de português colocados em escolas do ensino básico, tendo em conta inclusivamente o ajustamento dos temas à faixa etária dos alunos em estudo.²¹

Após a aplicação dos estímulos, foram obtidos 575 textos, produzidos por 287 alunos, distribuídos nos três anos de escolaridade definidos. Os alunos que produziram apenas um dos textos foram excluídos. Também, para participar deste estudo, somente os alunos nascidos em Portugal, falantes de português europeu e monolíngues, oriundos de famílias também monolíngues de português europeu, seriam considerados efetivamente. Para obter estas informações, no dia da aplicação do primeiro estímulo, foi fornecido a todos os alunos um questionário (Anexo 9.3) com perguntas sobre o local de nascimento e sobre a(s) língua(s) que se falava(m) em casa. Deste modo, seria possível excluir do estudo todos os participantes não-monolíngues ou falantes nativos de outras línguas que não o português. Em complementaridade, outras perguntas, entendidas como variáveis potenciais, constavam do questionário (sexo (masculino ou feminino); idade; e média a português no ano anterior).

3.3. Compilação do corpus

²⁰ Apesar de, nas provas finais do ensino básico e nos exames nacionais do ensino secundário, ser uma prática solicitar dos alunos a produção de textos limitada pelo número de palavras, as provas de aferição do ensino básico, diferentemente, costumam utilizar o número de linhas como critério limitador da produção. É esta a estratégia aqui adotada, já que o mesmo estímulo seria aplicado a alunos de escolaridade diferenciada, inclusive do 5.º ano.

²¹ Aos três professores, estou-lhes muito grato pelas valiosas contribuições. Assumo integralmente qualquer falha na elaboração das tarefas, como, por exemplo, a ausência do artigo antes de “teu (tua) melhor amigo (a)”, detetada após a aplicação aos alunos. Atribuo-a a um erro de digitação. Se necessário, toma-se esta situação como um viés estatístico.

3.3.1. Transcrição e armazenamento

Após a coleta total dos textos, iniciou-se a fase de transcrição e armazenamento dos textos, passando-se estes do formato manuscrito para o formato digital. Nesta fase, adotou-se um conjunto de procedimentos que garantisse como resultado, a um só tempo, uma coleção de textos que representasse o mais fielmente possível o texto produzido em sala de aula, sendo a referência para eventuais consultas ao longo do processo investigativo, e outra coleção de textos que pudesse ser submetida a um tratamento informático, orientada para o escopo desta tese, de que não fazem parte os habituais desvios de natureza gráfica, tais como pontuação, acentuação, capitalização ou indentação, componentes do texto a que Polio (2000: 102) chama de mecânicas.

Deste modo, foram criados, a partir de cada texto, dois ficheiros informáticos: o Ficheiro de Referência (FR) e o Ficheiro de Anotação (FA). Com o FR, que inclui a transcrição *ipsis verbis* dos textos manuscritos, pretende-se que sirva não apenas como fonte para eventuais consultas ao longo da fase analítica deste estudo, mas também como um recurso a investigadores e professores futuramente interessados no desenvolvimento linguístico da escrita escolar em Portugal. A transcrição dos textos manuscritos preservada no FR permite, portanto, que várias possibilidades de estudos se instaurem.

O Ficheiro de Anotação (FA) inclui a transcrição adaptada para fins de tratamento computacional e foram consideradas as seguintes regras básicas:

- preservar, sempre que possível, no FR e no FA a configuração gráfica relativa ao uso de maiúsculas e de pontuação apresentada nos textos manuscritos; nos casos de dúvida, a orientação geral foi a de beneficiar o escritor, transcrevendo-se com minúscula se nome comum ou maiúscula se nome próprio.

- manter no FR os pontos de marcação de abreviações, mas removê-los do FA, dado que podiam interferir na anotação morfossintática;

- manter no FR o recuo inicial de parágrafo de acordo com a configuração escolhida pelos escritores, eliminando-a do FA, já que a mancha gráfica seria irrelevante à investigação aqui empreendida;

- remover do FR e do FA outros casos relativos à representação gráfica, como sublinhados, linhas em branco (p. ex.: separação de títulos, saltos de linhas acidentais, etc.),

espaçamentos anormais entre palavras e *emoticons* ou quaisquer outros elementos não-verbais, o que se justifica por não serem factos linguísticos;

- preservar no FR e corrigir no FA os desvios de ortografia e acentuação gráfica, inclusive aparentemente acidentais, se não causassem problemas de interpretação;

- manter no FR e no FA as palavras ininteligíveis, mas descodificáveis pela grafia produzida pelo escritor.

- preservar no FR e no FA os termos estrangeiros (inclusive nomes próprios); no caso de a grafia destes termos estar incorreta, foi preservada no FR e corrigida no FA;

- recuperar, em sendo possível fazê-lo indubitavelmente, a omissão acidental de palavras; em caso contrário, manter-se como no original;

- preservar no FR e no FA o emprego de vulgarismos.

No processo de transcrição, questões relacionadas ao que Sampson (2005) chama de infelicidades estilísticas, ou seja, construções que seriam repudiadas ou mesmo tidas como incoerentes num texto produzido por um adulto escolarizado, mas que não se pode referir como agramaticais, tais como a repetição excessiva de palavras ou a coesão entre itens de género ou número díspares, foram preservadas tanto no FR como no FA pela razão óbvia de que esses fenómenos configuravam potenciais objetos de estudo desta tese. Incluiu-se ainda nesta categoria os desvios de flexão, concordância, regência, colocação pronominal e seleção lexical.

Como documento de referência quanto à ortografia e à acentuação gráfica, utilizaram-se os dicionários em linha Infopédia, da Porto Editora (www.infopedia.pt), e Dicionário Priberam da Língua Portuguesa, da Priberam Informática (<http://www.priberam.pt/DLPO/>).

Tantos os FA como os FR foram armazenados como texto plano (.txt) e nomeados com a seguinte composição: as letras iniciais do nome do aluno (p. ex.: abrm), o registo (p. ex.: 1 - narração ou 2 - argumentação), o ano escolar e a turma (p. ex.: 5c) e a escola de origem (p. ex.: mda). Todos os elementos de composição do nome foram separados por sinais de sublinhado (p. ex.: abrm_1_5c_mda), para facilitar a rápida discriminação quer visual, quer estatística na base de dados.

3.3.2. Ferramentas de anotação e consulta (*query*)

Como requisito necessário a todos os estágios de anotação dos itens de complexidade e da sua consulta (*query*), após a fase da segmentação, os textos foram anotados morfossintaticamente com o MBT (*memory-based tagger*) (Daelemans *et al.* 1996) e lematizados com o MBLEM (van den Bosch e Daelemans 1999), adaptados para o português (Généreux, M., I. Hendrickx, A. Mendes 2012).²²

Neste estudo, os indicadores de complexidade são regulados por duas dimensões do desenvolvimento linguístico – lexical e sintática –, razão por que foram precisas três ferramentas computacionais distintas. Para o trabalho de consulta das ocorrências de itens das dimensões lexicais, utilizou-se o programa IMS Open Corpus Workbench (CWB). Baseada no processador CQP (Centralized Query Processor), consiste numa coleção de ferramentas de código aberto que se destina a gerir e fazer consultas em *corpora* de grande extensão anotados morfossintaticamente. O CQP apresenta-se com muitos recursos, entre eles a possibilidade de uso de expressões regulares e *wild cards*, que permitem consultas bastante pormenorizadas, já que todas os itens dos textos estão associados a um atributo posicional (*pos*, *lemma* e *word*). Para a obtenção dos rácios relativos à diversidade lexical (medida *D*), recorreu-se à plataforma Clan (MacWhinney 2000), em particular ao programa *vocd*, que baseado em modelo de probabilidade, rastreia a potencial diminuição de rácios conforme o número de palavras num texto aumenta.

Para o trabalho de segmentação, anotação e consulta ao nível das unidades-t e orações, foi utilizado o programa UAM Corpus Tool (O'Donnell 2007), plataforma que permite um conjunto de procedimentos de tratamento de *corpora*, entre eles a segmentação e a anotação manuais de múltiplos textos nos seus diferentes níveis (documento, frase, oração, sintagmas e palavras) e também a exploração estatística de achados entre *subcorpora*. O UAM Corpus Tool também permite a criação de sistemas de etiquetas para a anotação orientada para fins específicos. Este recurso utiliza-se, em particular, para a anotação e a consulta de orações

²² Agradeço à Professora Amália Mendes a anotação morfossintática e a lematização nos textos do *corpus*.

dependentes. As ferramentas de anotação e consulta serão mais bem descritas nos capítulos pertinentes, em que se apresentam os procedimentos metodológicos específicos.

Para identificar os graus de relação entre variáveis independentes e dependentes e determinar as diferenças e semelhanças entre os grupos de escolaridade e os registos, foram aplicadas técnicas da estatística descritiva e inferencial (ou analítica), que são apropriadas para o domínio de testes de hipóteses e de significância e são bastante recorrentes em estudos baseados nas metodologias da linguística de *corpus*, isto porque a estatística inferencial não se presta à verificação de verdades, mas sim de probabilidades. Deste modo, um conjunto de testes estatísticos, no programa SPSS, a depender da normalidade dos dados, foi aplicado ao *corpus*: a) testes de correlação (Pearson ou Spearman); e b) análise de variância (ANOVA ou Kriskal Wallys) com *post-hoc* (de Tukey ou Bonferroni corrigido).

Aplicados os critérios de exclusão, participam efetivamente deste estudo 122 alunos, com 244 textos escritos, estando as informações gerais apresentadas na tabela abaixo, com informações gerais:²³

	5.º	7.º	10.º	TOTAL
N.º de participantes	26	46	50	122
Idade (M, DP)	10,19 (0,40)	12,33 (0,70)	15,16 (0,37)	–
Sexo				
feminino (%)	18 (69,2%)	25 (54,3%)	29 (58%)	72 (59%)
masculino (%)	8 (30,8%)	21 (45,7%)	21 (42%)	50 (41%)
Média a Português (M, DP) no ano anterior (escala 0-5)	3,92 (0,39)	3,46 (0,81)	3,84 (0,68)	–
N.º total de palavras	8.673	15.354	19.658	43.685
N.º total de textos	52	92	100	244

Tabela 1: Informações gerais

A diferença saliente de participantes do 5.º ano em relação aos outros anos reflete as dificuldades de acesso às escolas, maioritariamente impostas pelos próprios diretores. Neste

²³ Todos os valores absolutos das unidades linguísticas identificadas no estudo apresentam-se no Anexo 9.5.

sentido, configura-se a amostra aqui adotada como acidental ou de conveniência (não-randomizada), segundo classificação de Everitt e Skrondal (2010: 104). Neste tipo de amostra, a seleção de participantes é regulada por circunstâncias, que podem ser mais ou menos favoráveis à investigação. Embora os mesmos Everitt e Skrondal (*idem*) afirmem que, a partir de tal tipo de amostra, a representatividade da população geral seja improvável, aqui reitero a natureza descritiva da minha investigação, e não controlada, em que os tipos de amostras – randomizada ou não-randomizada – concorrem para a qualidade da representação. Quanto à diferença de participantes (e textos), procedeu-se à verificação da distribuição normal dos dados, em que foram feitos testes por variável (Shapiro-Wilk). Quando identificada a distribuição não-normal dos dados, foi feita a transformação (normalização) por meio da operação Log10.

Nos quatro capítulos seguintes, descrevem-se os indicadores de complexidade textual selecionados como variáveis dependentes tanto obtidas a partir do registo narrativo, como a partir do registo argumentativo, a serem correlacionadas com a variável independente ano escolar (5.º, 7.º e 10.º ano). Embora não haja fronteiras nítidas que individualizem os indicadores, porquanto eles estão inextricavelmente interrelacionados, como se explicita nas palavras de Ravid (2004: 340): “lexical items and syntactic constructions conspire to make a given piece of language more or less ‘complex’” –, aqui, unicamente para fins de classificação, os indicadores utilizados estão agrupados sob dois grandes rótulos: complexidade lexical e complexidade sintática. A escolha destes indicadores – e dos rótulos a eles aplicados – justifica-se pela sua estabilidade dentro da literatura sobre o desenvolvimento linguístico, quer em língua materna, quer em língua segunda.

Os indicadores de complexidade que, a seguir, se discutem, como sintetizam Wolfe-Quintero *et al.* (1998), são expressos matematicamente quer por frequências (médias – relativas ou absolutas) das unidades determinadas, quer por rácios.

4. Complexidade lexical

4.1. Introdução

Neste capítulo, apresenta-se o estudo de correlação entre a complexidade lexical, tratada como o conjunto de características de uso do léxico mensuráveis estatisticamente, e a progressão escolar nos textos escritos por crianças e adolescentes em idade escolar monolíngues de português europeu. Como afirmam Wray e Medwell (2006: 17), estudos de base empírica sobre o desenvolvimento do vocabulário nos anos finais da infância e na adolescência são surpreendentemente raros, se comparados com os estudos sobre a aquisição nos primeiros anos da infância. No entanto, Berman (2007: 348) lembra que é justamente o conhecimento vocabular a componente mais saliente e mais facilmente detetável do desenvolvimento linguístico de crianças e adolescentes em idade escolar. Neste período, não apenas um grande número de novas palavras é gradualmente incorporado no seu repertório, como também é adquirida a capacidade de expressar uma grande variedade, suportada, em especial, no uso de palavras de maior extensão e de menor recorrência no discurso quotidiano, advindas, por vezes, de registos especializados (Anglin 1993; Carlisle 2000).

Daller *et al.* (2007: 8) apresentam o conhecimento lexical de um aprendiz como um processo tridimensional, que compreende a extensão, que se refere à quantidade de palavras, a profundidade, que se refere à variedade do vocabulário, e a fluência, que se refere à velocidade de produção do vocabulário. Em linha semelhante, Read (2000: 197-198) propõe que a complexidade lexical²⁴ seja avaliada por quatro indicadores: a diversidade, que se mede pela razão entre a ocorrência de palavras diferentes e a totalidade de palavras num texto (também conhecida, no inglês, como *type-token ratio*); a densidade, que é a razão entre o número de palavras lexicais e o de palavras totais; a raridade, que se calcula pela razão entre

²⁴ Read (2000: 200 e seguintes) não usa a expressão “complexidade lexical”, mas “riqueza lexical” (no original, *lexical richness*). Por acreditar que a expressão “riqueza lexical” pode ser interpretada como pertencente a um domínio investigativo que ultrapassa os domínios desta tese, opto por usar a expressão “complexidade lexical”, mesmo que ainda não esteja devidamente estabilizada na literatura sobre o desenvolvimento linguístico em idade escolar. Pela mesma razão, opto por “raridade lexical”, em linha com Malvern *et al.* (2004: 5), em vez de “sofisticação lexical” (no original, *lexical sophistication*).

a quantidade de palavras de frequência incomum e a de palavras totais; e o número de erros, que se relaciona com falhas na ortografia, na flexão ou na derivação ou com a interferência de outras línguas.

Para a avaliação da complexidade lexical, terei em consideração dois dos indicadores referidos por Read (2000): a diversidade e a densidade, que são, pela sua natureza quantitativa, bastante mais adequados a um estudo baseado em *corpus*, além de já estarem estabilizados na literatura enquanto indicadores fiáveis de verificação do desenvolvimento linguístico do repertório lexical. A diversidade lexical consiste na razão entre a frequência de palavras diferentes e a frequência de itens totais (Templin 1957), na sua versão matematicamente corrigida, conhecida como medida *D* (Richards e Malvern 1997), e a densidade lexical consiste na razão entre a frequência de palavras lexicais e a frequência de palavras totais (Ure 1971; Halliday 2009), que aqui nomeio de medida *Z*.

Estabeleço como principal hipótese de trabalho que a diversidade e a densidade nos textos dos registos narrativo e argumentativo aumentam conforme o aluno avança nos anos escolares. Esta hipótese assenta na ideia comum (e sacralizada) de que, na escola, os processos de ensino e aprendizagem da escrita, e consequentemente do léxico, são graduais e ascendentes. Esta ideia reproduz-se no Programa de Português do Ensino Básico (Portugal 2009: 9-10), em que se afirma que “o processo de ensino e aprendizagem do idioma progride por patamares sucessivamente consolidados. De acordo com esta noção, a aprendizagem constitui um ‘movimento’ apoiado em aprendizagens anteriores”. Naturalmente, não se deduz de tal afirmação que o desenvolvimento seja sistemático e homogéneo para cada um dos alunos, mas sim que é possível distribuir este desenvolvimento em estádios discretos, mais ou menos ordenados, estando os alunos localizados em alguma parte do contínuo que vai de um estádio inicial a um estádio final. Como hipótese secundária, parto da presunção de que os textos do registo narrativo, em todos os anos escolares, apresentam maior diversidade e maior densidade que os textos do registo argumentativo. Justifica esta hipótese o facto de os textos argumentativos idealmente serem caracterizados por um léxico mais especializado (dependente do tópico, naturalmente), construído para fundamentar a existência de definições e conceitos, argumentos e contra-argumentos (favoráveis e contrários), que devem conduzir sequencialmente a conclusões. Um texto argumentativo serve, afinal, para influenciar as opiniões, as atitudes ou o comportamento de um interlocutor ou assistência, tornando credível ou aceitável um enunciado apoiado em outro argumento (Adam 1995:10).

4.2. Diversidade

A diversidade lexical é um indicador de desenvolvimento linguístico associado à quantificação da variação de palavras empregues num dado texto; ou seja, quanto maior a variação de palavras, maior a diversidade. Ransdell e Wengelin (2003, cf. McNamara *et al.* 2010: 57) sugerem que quanto maior é a diversidade lexical, mais patente é a competência linguística do falante/escritor. Por isso, pode-se concluir que um conhecimento limitado do vocabulário conduz à repetição e, conseqüentemente, reduz a complexidade de um texto (Stoddard 1990: 103). McCarthy e Jarvis (2010: 382) consideram que uma pequena taxa de diversidade lexical pode indicar a saturação temática num dado texto, ou seja, sem novas palavras, não há a introdução de novos temas.

É notável o reconhecimento deste indicador no cenário científico, o que se traduz na variedade de tópicos de investigação em que é utilizado e que vão além do desenvolvimento linguístico. Emprega-se este indicador, por exemplo, na estilística (Smith e Kelly 2002), para a verificação da autoria de textos históricos, na neuropatologia (Bucks *et al.* 2000), para a deteção das fases iniciais da doença de Alzheimer, ou na linguística forense (Colwell *et al.* 2002), para a identificação de testemunhos manipulados. Há ainda estudos, como o de Avent e Austermann (2003), que chegam mesmo a identificar uma correlação entre variação lexical e estatuto sócio-económico.

Sobre o desenvolvimento linguístico em idade escolar particularmente, vários estudos analisam a diversidade lexical produzida por crianças e adolescentes. São referências, por exemplo, os estudos de Berman e Verhoeven (2002), Stromqvist *et al.* (2002), Johansson (2008), Berman e Nir (2010), originados do projeto Spencer (Berman e Verhoeven 2002), cujo objetivo era examinar, em sete línguas diferentes (inglês, holandês, francês, islandês, hebreu, espanhol e sueco), as competências em língua materna não apenas de crianças e adolescentes, mas também de adultos, cobrindo uma faixa etária que se estende dos nove aos trinta anos. No contexto português, podem ser citados os trabalhos de Rodrigues (2008), que verifica a diversidade lexical num *corpus* de textos narrativos escritos por crianças monolíngues do 1.º ao 4.º ano do 1.º ciclo do ensino básico, e Costa (2010), que analisa a diversidade lexical em produções textuais de alunos do 4.º, 6.º e 9.º anos.

Tão variadas quanto as aplicações da diversidade lexical são as fórmulas para a sua aferição. A medida clássica para avaliar este indicador encontra-se em Templin (1957), para

quem a diversidade consiste na razão entre o número total de diferentes palavras, isto é, *types*, e o número total de palavras, *tokens*.²⁵ Esta medida é também conhecida simplesmente por TTR, forma abreviada da expressão em inglês *type-token ratio*²⁶. Para que um texto apresente uma alta taxa de diversidade lexical, o escritor deve recorrer a palavras variadas e evitar a repetição, motivo pelo qual a diversidade lexical é habitualmente associada à noção de produtividade (Wagner *et al.* 2011: 203). Os excertos abaixo exemplificam a aplicação da medida TTR (com *types* em itálico):

1. *Sim as redes sociais sim são importantes hoje em dia para a gente comunicar. Sim eu sou a favor porque as redes sociais são precisas tal como o Facebook e a gente fala por ele com a família e com os amigos. E primos e tias e etc... e muita mais gente.* (gpts_2_5c_mda)
2. *As redes sociais, hoje em dia, são um importante meio de comunicação, servindo também para o entretenimento das pessoas. O Windows Live Messenger permite-nos falar com amigos em tempo real, e para mim é uma das únicas maneiras que tenho de falar com pessoas que moram longe ou que já não vejo há muito tempo.* (ls2_2_10a_fdl)

O excerto (1), extraído de um texto argumentativo do 5.º ano, apresenta-se com 36 *types* distribuídos em 52 *tokens*, o que resulta numa taxa TTR de 0,69. No excerto (2), extraído de um texto argumentativo do 10.º ano, há um total de 46 *types* para 56 *tokens*, o que resulta numa taxa TTR de 0,82. Logo, o excerto argumentativo do 10.º ano, a considerar as taxas obtidas, é mais diverso no emprego do vocabulário, sustentando-se menos na repetição vocabular.

²⁵ No programa Clan, considera-se o item *a* um *type* diferente do item *à* ou o item *rede*, por exemplo, diferente do item *redes*.

²⁶ Para o português, Berber-Sardinha (2004: 94) propõe alternativamente as seguintes traduções para *type-token*: *forma-item* ou *vocabulo-ocorrência*. Por acreditar que nenhuma destas expressões já se tenha estabilizado plenamente na língua portuguesa, opto por utilizar as expressões em inglês.

Apesar da aparente eficiência da medida TTR para revelar a diversidade lexical de um texto, uma relação dependencial e progressiva entre esta medida como proposta por Templin ($TTR = V/N$, onde V corresponde a número de *types* e N a número de *tokens*) e a extensão do texto está reportada em McCarthy e Jarvis (2010: 381). Em termos práticos, à medida que um texto avança, mais palavras, ou itens²⁷, vão sendo incorporadas no mesmo, implicando, portanto, maiores quantidades de *tokens*, mas a taxa de crescimento de palavras diferentes diminui proporcionalmente; ou seja, quanto maior for um texto, menor será a TTR. Dito de outro modo, quanto mais se avança num texto, menos se pode extrair dele em relação ao seu vocabulário, o que se explica pela necessidade óbvia que um escritor tem de ser coesivo, de ser hábil em equilibrar o fluxo de informações dadas e informações novas, razão por que não pode abdicar totalmente dos dispositivos de coesão, entre os quais se encontra a repetição de vocábulos.

O problema associado à técnica TTR tem implicações na metodologia de uma investigação. Conscientes do problema, alguns investigadores, como Biber (1995), por exemplo, limitam a análise a porções reduzidas dos textos do *corpus*, ou limitam-na a um número específico de *tokens*. Em ambos os casos, não se consideram os textos na sua totalidade. Além disso, há autores, como Ertmer *et al.* (2002, cf. McCarthy e Jarvis 2010), que nem referem a problemática da extensão do texto, o que coloca em causa os resultados demonstrados.

Para contornar os inconvenientes impostos pela medida TTR, outras formas de cálculo são propostas, como fazem, por exemplo, Guiraud (1960) e Carroll (1964), que corrigem a

²⁷ É importante destacar que na aceção do termo *item* que está a ser empregue aqui sigo parcialmente Baker *et al.* (2006: 95), para quem *item* pode ser usado para designar qualquer termo pesquisável num *corpus*, quer se trate de uma palavra isolada, como *cão* ou *verão*, o que, nesta tese, bem se aplica à descrição da complexidade lexical, quer se trate de um grupo de palavras, como *já que*, *o cão*, *o meu melhor amigo*, o que bem se aplica para referir a complexidade sintática. Baker *et al.* (2006: 95) também incluem na categoria *item* construções tipicamente utilizadas para a consulta a *corpora*, como, por exemplo, *dog|cat* (de que resultará uma lista de todas as ocorrências dos itens *dog* e *cat* indiscriminadamente) ou *dog** (de que resultará uma lista de todas as ocorrências de itens iniciados por *dog*: *dogs*, *dogged*, *doggy*, etc.), em que recorrem a *wildcards* ou expressões regulares. A opção aqui é desconsiderar tais casos como *itens*.

relação dependencial entre a TTR e a extensão do texto pelo recurso quer à raiz quadrada, quer a algoritmos. No entanto, Durán *et al.* (2004: 221) garantem que essas reformulações matemáticas não superam o problema. Mais recentemente, Richards e Malvern (1997), a partir de proposições matemáticas de Brainerd (1982) e Sichel (1986), propõem a medida *D*, que, de acordo com McCarthy e Jarvis (2010), prevê a potencial redução da diversidade lexical em textos mais longos e, conseqüentemente, permite a comparação de textos de extensões distintas. A medida baseia-se num modelo de probabilidade para rastrear a diminuição da TTR conforme o número de *tokens* aumenta. Quanto maior for o valor *D*, maior será a diversidade lexical²⁸. Se aplicada aos mesmos excertos acima, (1) e (2), têm-se, respetivamente, as seguintes taxas: 36,49 e 96,08.

Como os textos escolares do *corpus* aqui investigado se apresentam com extensões variadas, opto pela medida *D*. São referências na aplicação desta medida como indicador do desenvolvimento lexical os estudos de Berman (2000) e de Stromqvist *et al.* (2001). Para o português, não encontrei estudos que utilizassem a medida *D*.

4.3. Densidade

O outro indicador de complexidade lexical considerado nos estudos sobre aquisição e desenvolvimento linguístico é a densidade lexical, assumida como um dos principais indicadores da trajetória do desenvolvimento rumo a uma escrita mais académica (Colombi 2002: 72), relacionando-se com a capacidade que um escritor possui de dispor a informação e condensá-la. Implica, portanto, os modos como a informação representada pelo léxico é incorporada na estrutura gramatical. Segue-se, assim, que um texto lexicalmente mais denso é um texto mais informativo, o que assenta no facto de serem os itens lexicais os portadores primários do conteúdo referencial (Ravid 2004: 346).

É importante salientar que a densidade (e naturalmente também a diversidade, já que se elabora sobre o léxico) se associa à capacidade cognitiva que crianças e jovens adquirem de usar a língua separada de contextos físicos imediatos, no que se convencionou chamar de língua descontextualizada, isto é, sem o suporte de contextos conversacionais (Snow

²⁸ Diferentemente da medida TTR, em que 1 é o máximo, a medida *D* não tem um limite preestabelecido.

1983, cf. Schleppegrell 2004: 8). Apesar de diversas críticas a que tal compreensão tem sido sujeita, a capacidade de produzir textos deslocados dos seus contextos reais, expressa, por exemplo, para referir o passado, antecipar o futuro ou especular sobre mundos possíveis alternativos à realidade, continua a ser bastante valorizada nos sistemas educacionais, sendo tipicamente referida como uma capacidade de ordem superior (MacLure 1999: 247). Por outras palavras, um texto com baixa densidade lexical configura-se como mais dêitico, mais situado no *aqui-agora* do discurso, sendo comumente um texto com características da modalidade falada; por oposição, um texto que apresenta taxas mais altas de densidade lexical estará mais próximo da modalidade escrita.

Como acontece com a diversidade lexical, também há diferentes formas de medição da densidade lexical, mas duas destas tornaram-se já tradicionais na literatura (Baker *et al.* 2006). A primeira proposta vem de Ure (1971) e Halliday (1987), que se fundamentam na relação entre os itens lexicais, ou palavras de conteúdo, e os itens gramaticais, ou palavras funcionais, que se estabelece num texto, expressando-a em forma de proporção. Posteriormente, Ure (1977) sugere uma reformulação à sua técnica: a proporção de itens lexicais pelos itens totais, o que é secundado por Stubbs (1996: 72), que acrescenta que a densidade deve ser expressa pela percentagem de ocorrências de itens lexicais, portanto multiplicando-se o resultado por cem. Tome-se, para a ilustração da aplicação da medida reformulada, os excertos anteriormente utilizados, reescritos em (3) e (4), com os itens lexicais em itálico:

3. Sim as *redes sociais* sim são *importantes* hoje em *dia* para a *gente comunicar*. Sim eu sou a *favor* porque as *redes sociais* são *precisas* tal como o *Facebook* e a *gente fala* por ele com a *família* e com os *amigos*. E *primos* e *tias* e etc... e muita mais gente. (gpts_2_5c_mda)
4. As *redes sociais*, hoje em *dia*, são um *importante meio* de *comunicação*, *servindo* também para o *entretenimento* das *pessoas*. O *Windows Live Messenger* *permite-nos falar* com *amigos* em *tempo real*, e para mim é uma das *únicas maneiras* que tenho de *falar* com *pessoas* que *moram longe* ou que já não *vejo há muito tempo*. (ls2_2_10a_fdl)

No exemplo (3), há um conjunto de 16 itens lexicais para um total de 52 itens (dos quais 36 são itens gramaticais), de que resulta uma taxa de densidade lexical, se considerada a fórmula itens lexicais/itens totais, de 0,31, de que se lê que 31% dos itens totais são lexi-

cais. Inversamente, do total de palavras empregues no excerto, 69% são palavras gramaticais. Quanto ao excerto (4), que tem um total de 56 palavras, das quais 25 são lexicais e 31 são gramaticais, tem-se um rácio de 0,45 item lexical, sendo possível afirmar que 45% dos itens totais equivalem a itens lexicais e que 55% são itens gramaticais.²⁹

A segunda proposta de medição da densidade lexical é também de Halliday (1987, 2005, 2009), para quem a densidade se mede pela frequência média de itens lexicais³⁰ por orações, excluindo-se as orações encaixadas (*embedded clauses*). O linguista sustenta que a oração é a unidade mais óbvia, mais natural para se avaliar a densidade, visto ser a maior unidade gramatical da língua, mas não exclui a possibilidade de se considerar outras unidades, apresentando recentemente uma definição mais abrangente (2009: 75): “the quantity of lexicalized information packed into a given unit in the grammar”. No entanto, Halliday afirma (2005: 168-9) que a aplicação do cálculo de densidade no escopo da oração, como ele propõe, é mais significativa quando na comparação da fala com a escrita, já que aquela, ainda segundo o linguista, tende a ser construída com mais orações do que esta. Por isto, não se utiliza esta forma de medição nesta tese, já que se trata de uma comparação entre textos unicamente da modalidade escrita. Acrescente-se que as medidas até este ponto referidas são convergentes em termos de resultados, como garantem Berman e Ravid (2009: 98), podendo resultar a sua aplicação em simultâneo ao mesmo conjunto de dados redundante.

Wolfe-Quintero *et al.* (1998: 105) listam várias outras técnicas de medição do uso do léxico, como, por exemplo, o rácio de nomes pelo total de *tokens* ou o rácio de verbos pelo total de *tokens*. Apesar de Wolfe-Quintero *et al.* (1998) referirem estas técnicas para a avaliação da diversidade, entendendo-as como mais adequadas à avaliação da densidade, já que envolvem a especialidade categorial do léxico, tornando-se, portanto, complementares às medidas clássicas de Ure (1971) e Halliday (1987). Deste modo, torna-se possível falar em densidade nominal (rácio de nomes pelos itens totais) ou em densidade verbal (rácio de

²⁹ É importante reiterar que o conceito de *item*, na complexidade lexical, equivale ao de palavra ortográfica, o que se expressa pela anotação do *item* ‘dia’, no excerto, como nome, e não como parte da locução adverbial de tempo.

³⁰ Em 2009, Halliday trata como sinónimos *itens lexicais* e *lexemas*.

verbos pelo total de itens). Neste trabalho, portanto, utiliza-se primeiramente a medição clássica (Ure 1971, Halliday 1985, Johansson 2009) – o rácio de itens lexicais (e as especialidades) pelo total de itens, expressa em forma percentual, e secundariamente, aplicam-se as medidas por especialidade (densidade por especialidade de classes de palavras).

As discussões sobre como calcular a densidade não se têm centrado somente em questões de natureza matemática, mas também em questões de natureza linguística. Dizem respeito à diferenciação entre itens lexicais e itens gramaticais. Ure (1971: 445), por exemplo, inclui como itens lexicais, ou itens com propriedades lexicais – em oposição aos itens com propriedades gramaticais – todos os nomes, verbos e adjetivos. Strömquist *et al.* (2001: 48) admitem como itens lexicais nomes, adjetivos e verbos, mas excluem os verbos auxiliares e os verbos copulativos. Há ainda autores que optam por descrever as ocorrências especializadas de itens lexicais, como fazem Ravid e Berman (2006), que avaliam as ocorrências de nomes num grupo de textos falados e escritos em registos narrativos e não narrativos.

Como observa Halliday (1989: 61), entre itens lexicais e itens gramaticais existe uma relação escalar, não existindo uma divisão nítida entre uns e outros, mas é possível fazê-la em favor da coerência metodológica do trabalho. Nesta tese, consideram-se as recomendações de Mendes (2013: 258), que, aceitando também que há uma relação escalar entre itens lexicais e itens gramaticais, admite como itens mais próximos do pólo lexical os seguintes: nomes, adjetivos (incluindo numerais ordinais), verbos plenos e advérbios terminados em -mente.

Apresentados os fundamentos necessários à compreensão das medidas de diversidade e densidade lexical, procede-se à descrição dos procedimentos metodológicos adotados especificamente para a obtenção destas medidas.

4.4. Procedimentos metodológicos específicos

Para a caracterização da diversidade lexical, os textos, convertidos para o formato .CHA, foram inseridos na plataforma Clan³¹ (MacWhinney 2000), de que se obtiveram os valores por texto da medida *D* (programa *vocd*). Na plataforma CQP, após definir como

³¹ A plataforma CLAN, bem como o manual, estão gratuitamente disponíveis para *download* (<http://chilides.psy.cmu.edu/>).

contexto apenas uma palavra (set Context 1) e solicitar a exibição dos atributos posicionais ‘pos’ e ‘lemma’ (show +pos; show +lemma) por texto (set PrintStructures “text_id”), foi extraída uma lista de *tokens* por ano e por registo textual.³² Os dados, delimitados por tabulação, foram, de seguida, importados para o programa Numbers, onde foi possível configurar uma lista das ocorrências de *types* com as respetivas frequências.

Para a caracterização da densidade lexical, com base na mesma lista extraída da plataforma CQP e convertida em base de dados no Numbers, delimitaram-se as ocorrências, por texto, de todas as categorias POS, sendo possível, de seguida, listar as ocorrências de *tokens* pertencentes às classes lexical e gramatical, diferenciadamente. Como já referido, a classe lexical é constituída por nomes, verbos plenos, adjetivos e advérbios terminados em -mente (Mendes 2013). No CQP, a classe ‘nome’ foi selecionada a partir das seguintes etiquetas: CN (nome comum), PNM (nome próprio), MTH (mês), WD (dia da semana) e STT (títulos sociais – como ‘Presidente’, ‘dr.’, ‘prof.’, etc.). Na composição da categoria ‘verbo pleno’, incluíram-se as etiquetas V (verbo), GER (gerúndio), INF (infinitivo), PPA (particípio não constituinte de tempo composto), excluindo-se as categorias GERAUX (gerúndio como verbo auxiliar), INFAUX (infinitivo como verbo auxiliar) e PPT (particípio em tempos compostos).³³ Também foram excluídos os verbos de ligação, ou verbos copulativos. Em virtude de o anotador MBT (Daelemans *et al.* 1996) não ser treinado para identificar, em português, verbos de ligação, recorreu-se ao anotador *online* Palavras (Bick 2000)³⁴. Os valores absolutos de *tokens*, *types*, *tokens* lexicais, *tokens* gramaticais estão apresentados no Anexo 9.5.

Considera-se, para todos os efeitos estatísticos, o ano escolar (5.º, 7.º e 10.º) combinado com o registo (narrativo ou argumentativo) como variável independente, já que a se-

³² Por não se tratar de uma preocupação desta tese, foram eliminadas as ocorrências de dígitos (DGT e DGRTR) e pontuação (PNT).

³³ Para mais informações, consulte-se o seguinte manual: “Manual for the CRPC CQPweb interface” (Généreux *et al.* 2011), disponível em: http://alfclul.clul.ul.pt/CQPweb/doc/CQPweb_manual.v1.pdf.

³⁴ Disponível em: <http://beta.visl.sdu.dk/visl/pt/parsing/automatic/parse.php>

leção vocabular é sensível ao registo (Johnson e Johnson 1999: 151). Como variável dependente, para avaliar a diversidade, consideraram-se os valores da medida *D* e, para avaliar a densidade, os valores da razão entre itens lexicais e itens totais, que aqui nomeio de medida *Z*. Dois testes estatísticos, no programa SPSS, foram aplicados aos valores obtidos: a) testes de correlação (Pearson) e b) análise de variância (ANOVA a um fator), com *post-hoc* de Tukey.

Antes da aplicação destes testes estatísticos, os valores das medidas *D* e *Z* foram verificados quanto à sua distribuição, se normal ou não, com base no teste de Shapiro-Wilk e na inspeção visual dos histogramas resultantes. A normalidade dos dados por este teste mede-se por um valor de $p \geq 0,05$. Constatou-se que os valores constituintes da medida *D* e os valores da medida *Z*, tanto no registo narrativo como no registo argumentativo, apresentavam uma distribuição anormal, sendo normalizados pela operação Log10. É sobre os dados transformados que se aplicam os testes estatísticos.

4.5. Estudo de correlação

Na Tabela 2, abaixo, apresentam-se, divididos por *registo* (narrativo e argumentativo), os valores médios de ocorrência de palavras totais (*tok*) e de palavras diferentes (*typ*), com respetivos desvios-padrão, identificados nos textos sob estudo. São estes os constituintes fundamentais da medida *D*, cujos valores médios, também diferenciados por registo, se apresentam mais abaixo, no Gráfico 1, seguindo-se a apresentação e discussão dos resultados dos testes estatísticos.

Registo	Ano	N	tok (DP)		typ(DP)	
Narrativo	5.º	26	179,12	45,14	101,62	19,53
	7.º	46	180,54	37,32	104,50	16,89
	10.º	50	207,62	48,40	129,20	25,25
Argumentativo	5.º	26	154,46	39,26	88,58	17,78
	7.º	46	153,24	38,26	91,61	20,80
	10.º	50	185,54	32,11	113,56	18,15

Tabela 2: Frequência média (e desvio-padrão), em números absolutos, de ocorrência de palavras totais (*tok*) e palavras diferentes (*typ*) identificada nos registos narrativo e argumentativo.

No registo narrativo, tanto o número médio de *tokens* como de *types* sobe do 5.º para o 7.º ano e deste para o 10.º ano. No registo argumentativo, o número médio de *tokens* decai ligeiramente do 5.º ao 7.º ano, mas, deste até ao 10.º ano, as médias elevam-se expressivamente. Os *types*, neste registo, no entanto, apresentam-se sempre ascendentes do 5.º ao 7.º e deste ao 10.º ano.

Quanto à medida *D* identificada no registo narrativo, os valores médios crescem de ano a ano. No 5.º ano, o valor médio obtido é de 78,27 (17,47³⁵); no 7.º ano, o valor médio é de 83,35 (27,88); no 10.º ano, o valor médio é de 109,87 (24,25). Quanto ao registo argumentativo, no 5.º ano, o valor médio de *D* é 74,44 (19,34); no 7.º ano, o valor médio é de 83,94 (22,12); no 10.º ano, o valor médio é de 100,76 (22,62). No Gráfico 1, é possível ilustrar o movimento de elevação da medida *D* do 5.º para o 7.º e deste para o 10.º ano, em ambos os registos, sendo a mudança aparentemente mais expressiva apenas a partir do 7.º ano:

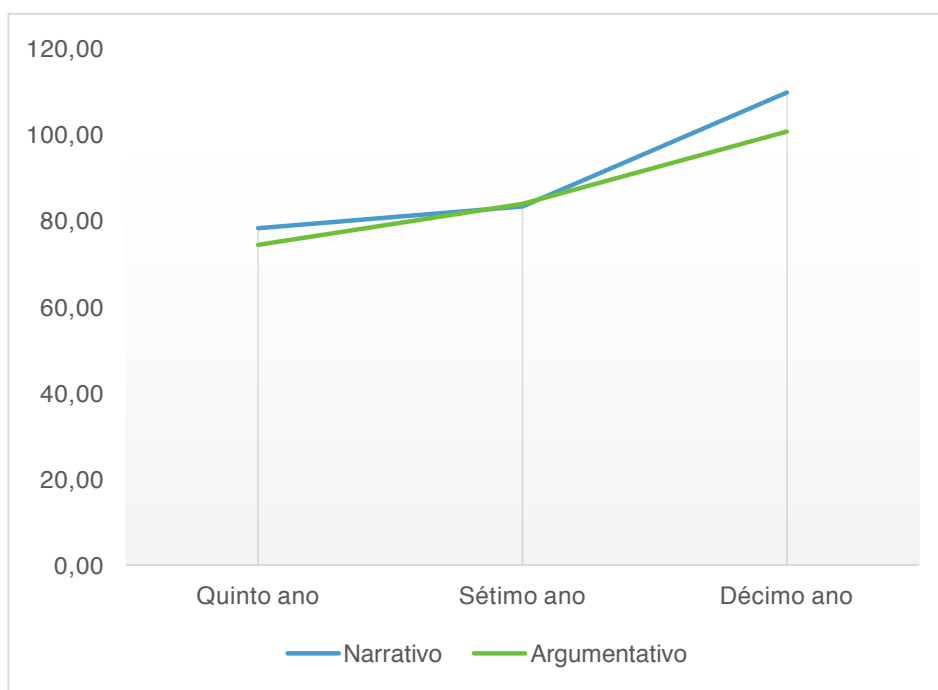


Gráfico 1: Distribuição dos valores médios da medida *D*, em números absolutos, nos registos narrativo e argumentativo ao longo dos anos escolares.

³⁵ Desvio-padrão.

Os testes de correlação indicam que, nos textos narrativos, há uma correlação positiva moderada³⁶ ($r=0,506$; $p=0,000^{37}$) entre a medida *D* e a progressão escolar. Na análise da variância, corrobora-se que há uma diferença significativa entre os grupos estudados ($F(22,924)$; $p=0,000^{38}$), mas o teste *post-hoc* de Tukey indica que somente há diferenças significativas entre o 5.º e o 10.º ano ($p=0,000$) e entre o 7.º e o 10.º ($p=0,000$), não havendo diferença significativa entre o 5.º e o 7.º ano ($p=0,780$). Quanto aos textos de registo argumentativo, há uma correlação positiva também moderada ($r=0,453$; $p=0,000$) entre a medida *D* e a progressão escolar. Na análise da variância, observa-se uma diferença significativa entre os anos escolares ($F(15,358)$; $p=0,000$). Em termos específicos, o teste *post-hoc* de Tukey revela que há diferenças significativas tanto entre o 5.º e o 10.º ano ($p=0,000$), como há também entre o 7.º e o 10.º ano ($p=0,001$), mas não há diferenças significativas entre o 5.º e o 7.º ano ($p=0,141$).

Na Tabela 3, abaixo, apresentam-se os valores médios de ocorrências de palavras lexicais (*lex*) e de palavras gramaticais (*gram*), acompanhados dos respectivos desvios-padrão, enquanto itens constituintes necessários à obtenção da medida *Z*:

Registo	Ano	N	lex (DP)		gram (DP)	
Narrativo	5.º	26	79,92	22,21	99,19	25,04
	7.º	46	78,63	15,27	101,91	25,19
	10.º	50	90,90	21,91	116,72	29,14
Argumentativo	5.º	26	69,23	17,17	85,23	24,75
	7.º	46	71,85	18,46	81,39	21,35
	10.º	50	86,96	16,24	98,58	18,08

Tabela 3: Frequência média (e desvio-padrão), em números absolutos, de itens lexicais (*lex*) e de itens gramaticais (*gram*) identificados nos registos narrativo e argumentativo.

³⁶ Segundo Dancey e Reidy (2004), a força da correlação avalia-se pelos seguintes valores de coeficiente (Pearson): coeficiente 1 = correlação perfeita; coeficiente entre 0,7 e 0,9 = correlação forte; coeficiente entre 0,4 e 0,6 = correlação moderada; coeficiente entre 0,1 e 0,3 = correlação fraca; e coeficiente 0 = correlação nula.

³⁷ Correlação significativa no nível 0,01 (2 extremidades).

³⁸ Diferença significativa no nível $<0,05$.

No registo narrativo, o número médio de palavras lexicais diminui do 5.º para o 7.º ano e eleva-se deste para o 10.º ano, enquanto o número médio de palavras gramaticais cresce linearmente de ano a ano. No registo argumentativo, acontece um processo contrário, em que o número médio de palavras lexicais cresce de ano a ano, mas o número médio de palavras gramaticais decai do 5.º para o 7.º ano, elevando-se deste para o 10.º ano.

Quanto à medida *Z* identificada no registo narrativo, os valores percentuais médios decrescem do 5.º para o 7.º ano, mantendo-se aparentemente estáveis até ao 10.º ano. No 5.º ano, o valor médio obtido é de 44,47 (4,10); no 7.º ano (4,93), o valor médio é de 43,93; e, no 10.º ano, o valor médio é de 43,94 (3,80). Quanto ao registo argumentativo, o movimento do valor percentual médio da medida *Z* é de ascensão do 5.º ao 7.º ano, e de ligeira diminuição deste ao 10.º ano. Assim, no 5.º ano, o valor médio de *Z* é 45,16 (4,71); no 7.º ano, o valor médio é de 47,00 (3,62); e, no 10.º ano, o valor médio é de 46,85 (3,23). No Gráfico 2, é possível ilustrar o movimento dos valores médios da medida *Z* aqui apresentados:

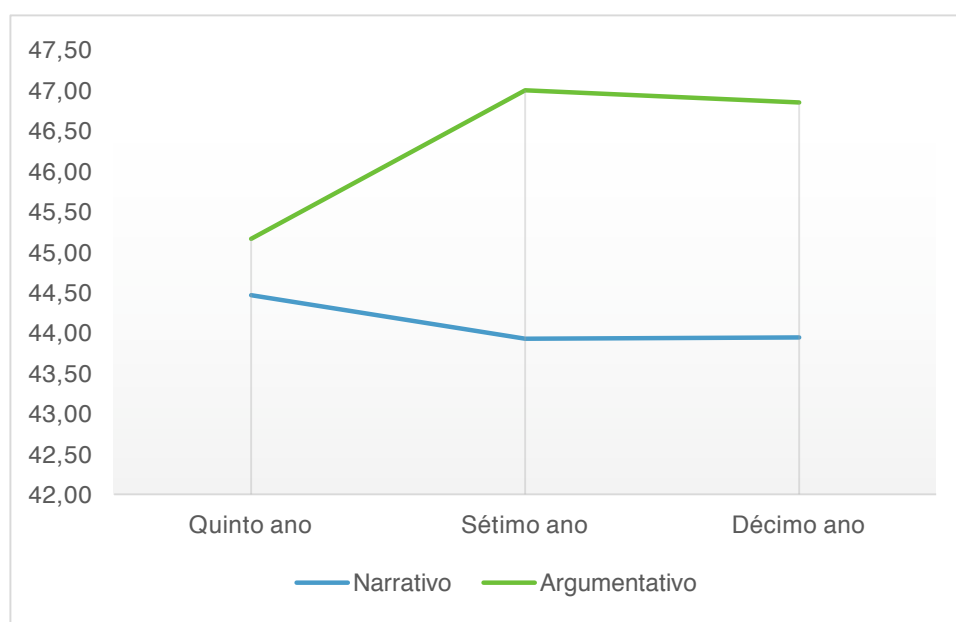


Gráfico 2: Distribuição dos valores médios da medida *Z*, em valores percentuais, nos registos narrativo e argumentativo ao longo dos anos escolares.

O teste estatístico, quanto aos textos narrativos, indica que não há correlação ($r=-0,031$; $p=0,736$) entre a medida *Z* e a progressão escolar. Quer isto dizer que, independentemente do estágio em que se encontra o aluno na escola, a proporção de itens lexicais em relação aos itens totais mantém-se inalterada. Quanto aos textos de registo argumentativo, os testes de correlação indicam que também não existe correlação ($r=0,142$; $p=0,120$) entre a medida *Z* e a progressão escolar, verificando-se situação semelhante ao que acontece nos textos de registo narrativo.

Na tentativa de compreender a ausência de correlações entre a medida *Z* e a progressão nos anos escolares, torna-se importante observar como se comportam individualmente as classes de palavras que constituem a categoria *lex*, o que se ilustra, no Gráfico 3 e no Gráfico 4, nos textos narrativos e nos textos argumentativos, respetivamente. Os valores apresentados são percentuais em relação ao total de ocorrências de *tokens*:

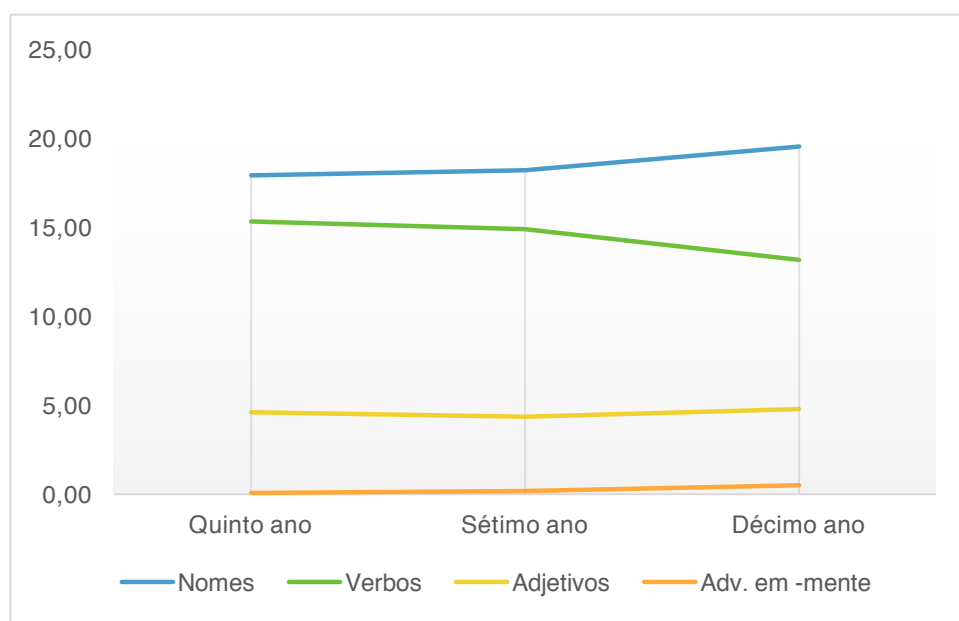


Gráfico 3: Distribuição dos valores médios de ocorrências, em percentual, de nomes comuns, verbos plenos, adjetivos e advérbios terminados em -mente no registo narrativo ao longo dos anos escolares.

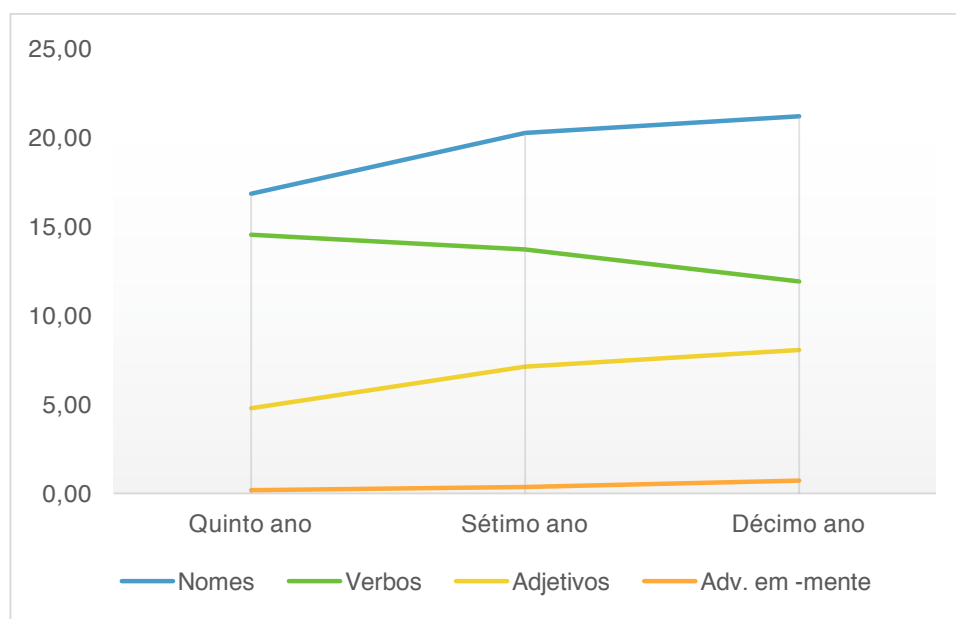


Gráfico 4: Distribuição dos valores médios de ocorrências, em percentual, de nomes comuns, verbos plenos, adjetivos e advérbios terminados em -mente no registo argumentativo ao longo dos anos escolares

O Gráfico 3, que trata das ocorrências das classes de palavras lexicais nos textos narrativos, permite ilustrar a ausência de correlação entre a medida Z e a progressão escolar, em que as percentagens de nomes, verbos, adjetivos e advérbios terminados em -mente mostram-se de uso homogêneo de ano para ano. A partir do Gráfico 4, que descreve esta medida nos textos argumentativos, pode-se inferir que a ausência de correlação entre a densidade lexical e a progressão escolar justifica-se nos valores gradualmente opostos das ocorrências de nomes e verbos, que, a partir do 7.º ano, se distanciam. No 5.º ano, os verbos são mais recorrentes do que os nomes; a partir do 7.º ano, passa-se a uma situação diferente, em que os nomes são mais acionados do que os verbos. Aparentemente, a percentagem de advérbios terminados em -mente mantém-se também inalterada ao longo dos anos, mas a de adjetivos mostra-se crescente de ano para ano. Contudo, a leitura dos gráficos acima não permite perceber, com precisão matemática, como se comportam as classes lexicais ao longo da progressão escolar. Deste modo, realizaram-se os testes estatísticos diferenciadamente para cada classe a fim de se obter dados sobre densidades por especialização (nominal, verbal, adjetival e adverbial).

Sobre o emprego de nomes no registo narrativo, o teste de correlação aponta para uma correlação positiva fraca ($r=0,240$; $p=0,008$) entre a densidade nominal e a progressão es-

colar. A análise da variância confirma a existência de diferenças estatisticamente significativas entre os anos de escolaridade ($F(4,834)$; $p=0,010$), com o teste *post-hoc* de Tukey a mostrar que essas diferenças se localizam apenas entre o 7.º e o 10.º ano ($p=0,011$), não se manifestando nem entre o 5.º e o 7.º ano ($p=0,924$), nem entre o 5.º e o 10.º ($p=0,093$). A densidade adjetival e a progressão escolar mantêm uma correlação positiva fraca ($r=0,234$; $p=0,009$), corroborada pela análise da variância ($F(4,819)$; $p=0,010$). No entanto, o teste *post-hoc* indica que só há diferenças entre o 7.º e o 10.º ano ($p=0,010$), não existindo nem entre o 5.º e o 7.º ($p=0,875$), nem entre o 5.º e o 10.º ($p=0,113$). Ainda quanto aos textos narrativos, não se identificou correlação entre a ocorrência de verbos e a progressão escolar ($r=0,129$; $p=0,156$). Por fim, foi identificada uma correlação positiva moderada ($r=0,404$; $p=0,000$) entre a densidade adverbial (-mente) e a progressão escolar, corroborada pela análise da variância ($F(12,072)$; $p=0,000$). O teste de Tukey mostra que não há diferença significativa entre o 5.º e o 7.º ano ($p=0,580$), havendo diferenças entre o 5.º e o 10.º ano ($p=0,000$) e entre o 7.º e o 10.º ($p=0,000$).

Quanto aos textos de registo argumentativo, identificou-se uma correlação positiva moderada ($r=0,418$; $p=0,000$) entre a densidade nominal e a progressão escolar, confirmada pelas diferenças encontradas na análise da variância ($F(12,661)$; $p=0,000$). As diferenças entre os anos, pelo *post-hoc* de Tukey, veem-se entre o 5.º e o 10.º ano ($p=0,000$) e entre o 7.º e o 10.º ano ($p=0,001$), não se identificando diferenças entre o 5.º e o 7.º ($p=0,295$). A densidade adjetiva nos textos argumentativos e a progressão escolar mantêm entre si uma correlação positiva moderada ($r=0,524$; $p=0,000$). A análise de variância confirma a existência de diferenças ($F(22,270)$; $p=0,000$), localizando-se estas diferenças na comparação entre todos os anos: entre o 5.º e o 7.º ($p=0,009$), entre o 7.º e o 10.º ($p=0,000$) e entre o 5.º e o 10.º ($p=0,000$). Ainda quanto aos textos argumentativos, pela aplicação dos testes estatísticos, não há correlação entre a ocorrência de verbos e a progressão escolar ($r=0,012$; $p=0,893$). Entre a densidade adverbial (-mente) e a progressão escolar identificou-se uma correlação positiva fraca ($r=0,347$; $p=0,000$). A análise da variância confirma a diferença entre os grupos estudados ($F(8,143)$; $p=0,000$), existindo especificamente diferenças entre o 5.º e o 10.º ano ($p=0,001$) e entre o 7.º e o 10.º ano ($p=0,019$), não acontecendo o mesmo entre o 5.º e o 7.º ($p=0,303$).

4.6. Discussão

Os resultados da aplicação dos testes de correlação apontam para uma tendência de crescimento da diversidade lexical, contabilizada por meio da medida *D*, ao longo da progressão do 5.º para o 10.º ano da escolaridade, tanto nos textos narrativos como nos textos argumentativos escritos pelos alunos participantes deste estudo, o que confirma a primeira hipótese de trabalho, segundo a qual a diversidade lexical aumenta conforme o aluno avança no seu percurso escolar. Todavia, como se viu pela aplicação dos testes *post-hoc*, o crescimento não é constante entre os três anos estudados, não se revelando diferenças significativas na diversidade lexical entre o 5.º e o 7.º ano, nem no registo narrativo, nem no registo argumentativo. Quer isto dizer, em termos estritamente matemáticos, que a variação no uso do vocabulário identificada no 5.º ano se assemelha à variação identificada no 7.º ano, havendo mudanças apenas na comparação com a variação no 10.º ano da escolaridade, pelo que é possível afirmar que é em algum momento entre o 7.º e o 10.º ano dos grupos estudados que se inicia, de facto, um processo de maior diversificação do uso das palavras.

Em Berman e Verhoeven (2002), Stromqvist *et al.* (2002), Johansson (2008) e Berman e Nir (2010), reporta-se uma situação semelhante. Segundo estes estudos, há um significativo desenvolvimento da diversidade lexical, também examinada com base na medida *D*, apenas entre os treze e os dezassete anos, sendo o desenvolvimento pouco significativo dos nove aos doze e dos dezassete aos trinta anos. Dois outros estudos, vistos em complementação, corroboram as conclusões dos estudos anteriores. Por um lado, Malvern *et al.* (2004), analisando textos produzidos por crianças e adolescentes em três diferentes grupos etários (7, 11 e 14 anos), notam que os resultados dos textos dos escritores mais velhos apresentam os resultados mais expressivos, sendo inexpressiva a diferença entre os dois grupos mais novos. Por outro lado, Crossley *et al.* (2011) identificam em textos escritos também por três grupos etários (14-15 anos, 16-17 anos e primeiranistas universitários) diferenças estatísticas significativas entre todos os grupos. Para efeitos de comparação com os resultados aqui encontrados, vale a pena lembrar que, no percurso normal da escolaridade portuguesa, aos treze anos está-se no 8.º ano.

Para o português europeu, não foram identificados estudos que tratassem da diversidade lexical pela medida *D*, que aqui se considera, ou que fizessem a mesma abordagem quantitativa. Apesar disto, o estudo de Costa (2010) pode ser tomado como uma referência. Com o objetivo geral de examinar o uso de conetores de valor concessivo em diferentes tipos de produções escritas de alunos do quarto, sexto e nono anos da escolaridade básica,

a investigadora (2010: 137-8) considera, enquanto estratégia complementar de caracterização quantitativa do *corpus*, a frequência média de *types* por texto, a partir do que identifica, do quarto para o nono ano, um aumento no uso de *types*. Costa observa que há um crescimento da diversidade lexical ao longo da progressão escolar, mas que, entre o quarto e o sexto ano, não se vê esse crescimento especificamente.³⁹

Com base na síntese acima, é possível afirmar, em linha com Stromqvist *et al.* (2002: 53), que é entre os treze e os dezassete anos que há um salto expressivo na diversidade lexical da escrita escolar, pelo que se deve acautelar, como lembra Johansson (2009: 77), que as medidas de descrição da diversidade lexical entre grupos de pequenas diferenças de idade não devem ser usados sem a complementação de outras medidas.

Ainda quanto aos resultados dos testes estatísticos aplicados ao *corpus* considerado, viu-se que a força da correlação entre a medida *D* e cada registo isoladamente confirma a segunda hipótese de trabalho, segundo a qual os textos do registo narrativo apresentam maior diversidade lexical que os textos do registo argumentativo. Como se viu, a força da correlação nos textos narrativos é mais expressiva do que nos textos argumentativos. Conforme Ravid (2004: 339), muitos estudos afirmam que, desde os cinco ou seis anos de idade (antes do período escolar, portanto), as crianças já possuem conhecimentos esquemáticos e linguísticos bem estabelecidos sobre a produção de textos narrativos, não se podendo afirmar o mesmo sobre os textos argumentativos, que só serão plenamente dominados na adolescência. Com isto, não se quer dizer que crianças e adolescentes não saibam distinguir um registo do outro, mas sim que lhes falta, até à adolescência, a familiaridade com o conjunto

³⁹ Por ser uma medição relativamente fácil, verificaram-se estatisticamente, no *corpus* aqui estudado, os valores médios de *types* por texto. A aplicação dos testes, revela que, nos textos narrativos, há uma correlação positiva moderada entre a frequência de *types* ($r=0,495$; $p=0,000$) e a progressão escolar. Situação semelhantemente foi identificada nos textos argumentativos ($r=0,496$; $p=0,000$). A análise da variância confirma que há diferenças entre os grupos estudados, tanto nos textos narrativos ($F(21,854)$; $p=0,000$), como nos textos argumentativos ($F(21,638)$; $p=0,000$). O teste *post-hoc* de Tukey mostra que, nos textos narrativos, só há diferenças significativas entre o 5.º e o 10.º ano ($p=0,000$) e entre o 7.º e o 10.º ano ($p=0,000$), não havendo diferença entre o 5.º e o 7.º ($p=0,845$). Nos textos argumentativos, do mesmo modo, só há diferenças significativas entre o 5.º e o 10.º ano ($0,000$) e entre o 7.º e o 10.º ano ($0,000$), não havendo entre o 5.º e o 7.º ano ($0,795$).

de conhecimentos linguísticos e retóricos que caracterizam o registo argumentativo, cujo foco é a apresentação de ideias e conceitos com base em argumentos e contra-argumentos (Berman e Slobin 1994, cf. Ravid 2004: 339).

Quanto à correlação entre a densidade lexical, contabilizada por meio da medida *Z* (razão entre a ocorrência de itens lexicais e o total de itens, multiplicada por cem), e a progressão do 5.º ao 10.º ano da escolaridade, tanto nos textos narrativos como nos argumentativos, como se viu, não há movimentos de mudança significativos. Este resultado contraria a expectativa de que, na progressão escolar, há um adensamento lexical da escrita e, conseqüentemente, refuta a primeira hipótese de trabalho, segundo a qual os textos dos alunos mais velhos são mais densos lexicalmente do que os textos dos alunos mais novos, pensamento que se vê, por exemplo, em Schleppegrell (2004: 108), que afirma que quanto mais se avança nos anos escolares, mais os alunos se tornam capazes de condensar informação, através de itens lexicais. Este resultado também anula, em princípio, uma possível percepção de que a escrita escolar pode ser tratada como uma faceta da escrita académica, como considera Jisa (2004: 135). Contudo, Biber (2006: 14) destaca que a escrita académica se caracteriza pela prevalência de nomes e pela ocorrência menos saliente de verbos, resultados que encontram eco nos textos aqui estudados quando se considera a densidade por classe, sendo identificadas correlações positivas, em ambos os registos, entre o emprego de nomes, adjetivos e advérbios em -mente e a progressão escolar, e nenhuma correlação entre o emprego de verbos e a progressão escolar.

Berman e Ravid (2009), num estudo experimental que considera participantes em idade escolar (dos nove aos dezanove anos) monolíngues de hebreu ou de inglês, relatam, por exemplo, que a taxa de densidade lexical, em particular de adjetivos, cresce significativamente em função da idade e em função do registo. Mais especificamente, os textos produzidos pelos participantes mais velhos, em idade escolar mais avançada, apresentam maiores taxas de densidade do que os textos dos mais novos. Ravid (2004) avalia a densidade global e específica de itens lexicais por texto num estudo que compara a complexidade linguística em textos expositivos e narrativos escritos por alunos hebreus monolíngues situados no 4.º, 7.º e 11.º anos de escolaridade e constata que a densidade nominal cresce significativamente de ano a ano.

Os resultados dos testes estatísticos aplicados ao *corpus* aqui investigado, em que não se identificou correlação nenhuma, refutam consequentemente a segunda hipótese de trabalho, a de que a densidade lexical, considerada holisticamente, é maior nos textos narrativos, já que, como alguns estudos sugerem (Bereiter e Scardamalia 1982; Christie 1986), os alunos, ao longo da sua vida escolar, experimentam muitas dificuldades na escrita de textos não-narrativos, tornando-se estes tipos de textos grandes desafios, inclusivamente na seleção lexical, pois, mesmo sendo dependentes de um dado tema, habitualmente requerem um léxico mais especializado do que os textos narrativos. A refutação da hipótese secundária, se considerada a densidade por classe, mantém-se, desta feita não pela ausência de correlação, mas sim, como se verificou, pela força da correlação da densidade nominal, verbal e adjetival no registo narrativo, que é sempre menos expressiva do que a força no registo argumentativo. Semelhantemente, os resultados de Ravid (2004) apontam para um crescimento mais acentuado da densidade nominal e adjetival nos textos argumentativos. Berman e Ravid (2009), sobre a densidade adjetival, concluem que o crescimento é mais significativo nos textos expositivos do que o que se encontra nos textos narrativos. Para o português europeu, não foram identificados estudos sobre o desenvolvimento linguístico que se propusessem aferir a densidade lexical em textos escolares pela mesma métrica que aqui se utiliza.

Os resultados de Ravid (2004), Berman e Ravid (2009), em complementação com o que aqui se identificou, parecem desafiar, pelo menos quanto à seleção lexical, a conceção, lembrada por Ravid (2004: 351), de que os textos narrativos produzidos por crianças e jovens em idade escolar, por serem adquiridos ainda nos anos iniciais da escolaridade e por serem aparentemente mais bem escritos, se mostram mais complexos linguisticamente do que textos de outros registos. Isto não quer dizer que os textos argumentativos escritos pelos alunos portugueses sejam melhores ou mais bem escritos do que as suas contrapartes narrativas. Pode, ao contrário, apontar para o facto de estes alunos, quando confrontados com tarefas que lhes exigem uma ação verbal distinta daquela a que estão habituados, se esforçarem mais, demonstrando consciência sobre os mecanismos de produção dos registos e das suas especificidades lexicais.

No registo argumentativo, ainda quanto aos resultados da medição da densidade por classe, observou-se que, à exceção do adensamento do léxico nominal e adjetival, que se torna significativo já entre o 5.º e o 7.º ano (e consequentemente entre o 5.º e o 10.º), em

todos os outros casos de densidade por classe, as diferenças entre os grupos são significativas apenas entre o 7.º e o 10.º ou entre o 7.º e o 10.º. Nos textos de registo narrativo, todas as diferenças identificadas igualmente ocorrem apenas entre o 7.º e o 10.º ano (ou entre o 5.º e o 10.º).

4.7. Síntese

O estudo aqui reportado teve por objetivo central examinar a correlação entre indicadores de desenvolvimento da complexidade lexical, nomeadamente a diversidade e a densidade lexical, e a progressão nos anos escolares em dois registos (narração e argumentação). É, portanto, um trabalho de natureza descritiva, que tenta cumprir o seu objetivo ao apontar para valores estatísticos que podem ser, com as devidas ressalvas, tomados como tendências do desenvolvimento linguístico ao longo da evolução na escola.

Como se constatou, há uma correlação positiva entre a diversidade lexical (proporção de palavras *types* por *tokens*) e a progressão escolar do 5.º ao 10.º ano da escolaridade básica em ambos os registos, ou seja, quanto mais se avança nos anos de escolaridade, ciclo a ciclo, mais variado se torna o vocabulário dos alunos. Esta correlação expressa-se mais fortemente no registo narrativo. No entanto, a correlação em ambos os registos caracteriza-se por não ser linear, já que não foram detetadas diferenças estatísticas significativas entre o 5.º e o 7.º ano na quase totalidade dos testes.

Quanto ao indicador de densidade lexical (proporção de palavras lexicais pelo total de palavras), não se identificou correlação, nem positiva nem negativa, em nenhum dos registos, ou seja, o fato de o aluno avançar nos anos de escolaridade não tem implicações no emprego de palavras de propriedades lexicais. Confrontou-se este resultado com os resultados de uma abordagem por classe, em que, no registo narrativo em particular, se identificaram correlações positivas entre nomes e advérbios terminados em *-mente* e o avanço nos anos de escolaridade, e, em oposição, identificou-se uma correlação nula na classe dos verbos. Nos textos argumentativos, os indicadores associados a nomes, adjetivos e advérbios terminados em *-mente* mantêm correlação negativa com a progressão escolar, excetuando-se o indicador associado ao verbo, que não mantém correlação com a progressão escolar. É importante salientar que as correlações mais fortes foram detetadas nos textos ar-

gumentativos. Além disso, tal como se viu quanto à diversidade, a correlação entre a densidade lexical por classe e os anos escolares também não é linear, não existindo, de acordo com a maioria dos testes, diferenças significativas entre o 5.º e o 7.º ano.

O facto de não haver diferenças significativas no desenvolvimento da diversidade lexical em textos escritos entre o 5.º e o 7.º ano, havendo apenas em relação ao 10.º ano, e também o facto de a densidade lexical globalmente não ser uma função da progressão escolar podem ser surpreendentes quando confrontados com a expectativa comum de que na escola de ano para ano há um crescimento progressivo das competências linguísticas dos alunos. Na verdade, trata-se do reflexo de um persistente problema relacionado com o desenvolvimento da escrita escolar, bem sintetizado na seguinte afirmação de Wilkinson *et al.* (1980: 2): “Development obviously takes place, but does not take place obviously”. Quer isto dizer que, apesar da certeza de que a escrita de um rapaz de 16 anos é provavelmente melhor do que a escrita de um rapaz de 11, ainda não se pode afirmar, com precisão, como esta melhoria, em termos desenvolvimentais, se realiza.

5. Complexidade sintática

5.1. Introdução

O objetivo deste capítulo é apresentar o estudo de correlação entre a complexidade sintática e a progressão escolar nos textos escritos por crianças e adolescentes em idade escolar monolíngues de português europeu. Entende-se por complexidade sintática o conjunto de padrões de uso de estruturas oracionais, medíveis estatisticamente.

Desde a década de 60, a complexidade sintática tem sido considerada em muitas investigações sobre o desenvolvimento linguístico de L1 de crianças e jovens (e de L2, por extensão). Estas investigações revelam-se expressivas particularmente nas décadas de 60, 70 e 80 (Harpin 1973, 1976; Loban 1963, 1976; O'Donnell *et al.* 1967; Perera 1984) e têm tido impacto não apenas no âmbito da investigação, mas também nas práticas pedagógicas. De facto, no seu trabalho seminal, Perera (1984: 4) sugere que se reconheça a importância da maturidade (ou imaturidade) linguística através da sintaxe, o que permitirá aos professores ajudar melhor os seus alunos. As recomendações curriculares de Loban (1986, cf. Schleppegrell 2004: 14) são até hoje referenciadas na elaboração de matrizes curriculares nos Estados Unidos.

Nos anos 90, porém, há uma considerável diminuição do número de estudos, quando a preocupação sobre os padrões de realização linguística em L1, de acordo com Biber *et al.* (2011: 6), cede espaço a estudos focalizados nos processos cognitivos da escrita, no pensamento crítico e nos contextos socioculturais de textos produzidos em ambiente escolar. A complexidade sintática passa a ser tópico recorrente apenas em pesquisas sobre a aquisição de L2. Para Scott (2004), o abandono da centralidade da sintaxe, refletido no silêncio de vários relatórios publicados por comissões oficiais dos Estados Unidos, teve um custo explícito, que se vê no enfraquecimento da literacia nacional, tanto na leitura como na escrita. Com o advento de recursos informáticos mais eficientes para o tratamento de grandes quantidades de texto, renova-se a curiosidade sobre o desenvolvimento linguístico em idade escolar, com McNamara *et al.* (2010), Beers e Nagy (2011), Berninger *et al.* (2011), Biber *et al.* (2011), entre outros, à frente desta renovação.

Dos estudos passados e recentes, pode-se concluir que a presença de certos comportamentos gramaticais implica um controlo crescente da língua. Hunt (1965: 5), referindo-se

à escrita de crianças e adolescentes, associa este tipo de comportamento à maturidade sintática, que se refere a um conjunto de características que são observadas com maior incidência em textos de escritores mais velhos e com fraca ou nenhuma ocorrência nos textos de escritores mais novos. Foster e Skehan (1996: 303) apontam para uma situação semelhante e salientam que o desenvolvimento gramatical do uso da língua consiste num processo crescente de elaboração, representado por uma grande variedade de padrões sintáticos. Assim entendida a complexidade sintática, rejeita-se a ideia de que ela possa corresponder à simples quantificação de estruturas indistintas. Reforça-se, pelo contrário, o fundamento de que a complexidade, em todas as suas dimensões, incluindo a sintática, é um fenómeno representativo da variação do uso que se faz da língua.

Embora a complexidade sintática se possa resumir a aspetos do nível frásico, e em especial a aspetos da subordinação, há algumas peculiaridades envolvidas no tratamento dos indicadores de complexidade sintática que merecem ser explicitadas, não apenas no que respeita aos conceitos e às técnicas aplicadas na sua medição, mas também no que respeita às unidades e às estruturas que fundamentam essas técnicas. A primeira peculiaridade refere-se às unidades que servem de base para a medição da variação da complexidade sintática. Wolfe-Quintero *et al.* (1998: 274) observam que os indicadores considerados na generalidade dos estudos sobre o desenvolvimento linguístico dividem as suas preferências entre a oração e a unidade-t, as quais tanto podem figurar isoladamente, sendo o foco a sua constituição interna, como em combinação umas com as outras. A oração, entendida como unidade sintática, é definida por Raposo *et al.* (2013: 314) como “uma sequência de palavras gramatical, que tem um conteúdo proposicional e como elemento nuclear um verbo”, definição que aqui se adota. A consideração da unidade *oração* tem como principal vantagem, enquanto unidade de análise, a fácil identificação na superfície textual escrita, além de ser uma unidade fiável tanto para uma análise de base sintática, como para uma análise de base semântica ou discursiva.

A unidade-t, proposta por Hunt (1965: 21), é a unidade constituída por uma oração independente e, quando for o caso, pela(s) oração(ões) que a ela se subordina(m). Se a unidade-t se constitui unicamente por uma oração independente, trata-se de uma unidade-t simples; ao contrário, se há uma ou mais orações subordinadas, é o caso de uma unidade-t multioracional. Portanto, a unidade-t é uma unidade autónoma das unidades que a antecedem ou se lhe seguem na superfície textual. Biber *et al.* 2007: 1069-70), sobre a unidade-c

(*c-unit*), proposta por Loban (1976), afirma que se configura como a menor unidade discursiva para fins de segmentação textual. Por analogia, pode-se dizer o mesmo da unidade-t, já que estas duas unidades se distinguem apenas pelo objeto a que se aplicam: aquela para a fala e esta para a escrita. Como já discutido no Capítulo 2.1, a escolha pela unidade-t é particularmente recorrente em estudos sobre a escrita escolar porque, segundo Hunt (1965: 20-21), permite a um analista contornar os inconvenientes que a pontuação de aprendiz lhe impõe.

Wolfe-Quintero *et al.* (1998: 274) referem, em paralelo com a unidade-t e a oração, a frase como uma unidade possível para analisar a complexidade sintática. No entanto, acredito que incluí-la como unidade de análise de textos escritos escolares pode não ser coerente com a unidade-t, que, afinal, lhe é uma alternativa. Além disto, como lembra Linell (2005: 169), o termo *frase* é regularmente usado de modos distintos, por vezes com um sentido vago ou ambivalente, confundindo-se com o termo *oração*.

Por fim, a segunda peculiaridade refere-se às técnicas de medição da complexidade sintática. Por ser já uma tradição a utilização de indicadores de complexidade sintática para examinar o desenvolvimento linguístico da L1 de crianças e jovens, há muitas técnicas de medição que têm sido propostas. Contudo, é possível sintetizá-las em duas tendências de procedimentos, pelo que proponho a seguinte classificação: o procedimento com base na extensão, ou quantitativo, e o procedimento com base na variação, ou qualitativo. O primeiro considera a ocorrência numérica dos constituintes de uma dada unidade linguística, sem diferenciar tipos específicos de ocorrência; o segundo procedimento considera a diversidade desses constituintes, avaliados por tipos específicos.

Com base nos dois procedimentos de medição (a seguir, descritos mais pormenorizadamente), estabeleço como primeira hipótese de trabalho que a extensão sintática se correlaciona positivamente com a progressão nos anos escolares em ambos os registros, ou seja, conforme o aluno progride na escolaridade, as orações e as unidades-t que produzem tornam-se mais robustas em termos de constituição vocabular e as unidades-t tornam-se mais robustas em termos de constituição oracional. Como consequência desta mudança, prognostico ainda que a variação também mantém uma correlação positiva com a progressão escolar, refletindo uma utilização mais diversificada dos quatro tipos (e dos seus subtipos) de orações subordinadas em oposição a um uso cada vez menos expressivo de estruturas oracionais simples ou coordenadas. Em conformidade com o que sintetiza Perera (1986:

496), estabeleço como hipóteses sobre a variação da complexidade sintática, por um lado, que as orações subordinadas completivas e adverbiais apresentam rácios de proporção de ocorrência regulares ao longo dos anos escolares, mas, por outro lado, as orações relativas e gradativas⁴⁰ têm taxas de ocorrência crescentes de ano a ano. É importante ressaltar, contudo, que o desenvolvimento da variação quanto aos tipos de subordinação não seja homogêneo, já que, como lembra Perera (1986: 496), estas orações têm diferentes propriedades gramaticais e funcionais, sendo muitas vezes dependentes do tópico e dos propósitos, e acréscimo, do registo, em que o texto pretende enquadrar-se. Enquanto prática já estabelecida no capítulo anterior, os casos específicos da realização das orações subordinadas que sejam considerados (numericamente) relevantes apresentam-se na discussão dos resultados.

Dada a natureza dos textos do registo argumentativo, que são tipicamente interpretados como mais exigentes ao escritor, tanto cognitiva como linguisticamente (Nippold 2004: 12), tem-se por hipótese secundária que, em ambas as dimensões de avaliação da complexidade sintática – a extensão das orações e das unidades-t e a variação no uso dos tipos de oração subordinada –, notar-se-ão mudanças mais significativas nos textos de registo narrativo do que nos textos de registo argumentativo. Esta hipótese contraria as assunções de Ravid (2004: 351), segundo a qual os textos argumentativos são mais ricos e mais complexos na sua expressão linguística do que os textos narrativos, sendo uma consequência dessa complexidade, como defendem Beers e Nagy (2011: 188), a identificação das diferenças ao nível sintático como mais saliente naquele tipo de textos do que neste. Trata-se de uma hipótese que se justifica numa perceção empírica de que a prática escolar, confirmada pelo Programa de Português do Ensino Básico, conduz os alunos a tornarem-se, ano a ano, mais familiarizados com os textos do registo narrativo, como se pode ver pela inserção deste tipo de textos desde os anos iniciais do primeiro ciclo da escolaridade e pela sua manutenção como uma necessidade de aprendizagem até, pelo menos, o 7.º ano, onde tais textos figuram para corroborar as “práticas de relato e reconto de experiências, de acontecimentos, de filmes vistos ou de livros lidos” (Portugal 2009: 173), enquanto conteúdo formalmente proposto para o ensino das habilidades de escrita, aparecendo os textos argumentativos como uma preocupação do ensino da escrita apenas a partir do segundo ciclo da escolaridade.

40 Perera (1986) considera as orações gradativas como um tipo de oração adverbial, diferentemente do que faço aqui.

As hipóteses acima colocadas têm a sua origem em orientações e práticas pedagógicas que comumente defendem que a aprendizagem do português, no plano sintático, deve ser organizada pela noção de complexidade crescente, como se pode ver no Programa de Português do Ensino Básico, em que se considera, por exemplo, a mudança da “frase simples à frase complexa” por meio da expansão do grupo nominal (2009: 33) ou a “articulação de frases complexas” (2009: 95) como estratégias para assegurar o domínio quer da expressão oral, quer do conhecimento explícito da língua. Assume-se como expectável que o uso da língua em produções textuais efetivas espelhe tais aquisições, sendo incorporadas gradualmente conforme a criança progride nos anos escolares, o que se confirma por estudos que asseguram que crianças e adolescentes nos anos finais da escolaridade produzem textos escritos com mais palavras e, interligadamente, com mais orações, e mais variadas quanto à funcionalidade semântica, do que os alunos dos anos iniciais (Malvern *et al.* 2004; Ravid 2005; Nippold *et al.* 2005; Nippold *et al.* 2007; Beers e Nagy 2011).

5.2. Extensão

O procedimento de medição da complexidade sintática com base na extensão fundamenta-se na frequência de uma dada unidade dentro de outra, que lhe é superior, portanto. Tipicamente, obtém-se a extensão pela frequência média de palavras por frase (ou enunciado), o que se conhece pela sigla MLUw (do inglês *mean length of utterance in words*), que é uma adaptação de uma medida proposta por Brown (1973) para quantificar, de um ponto de vista clínico, a extensão de enunciados em morfemas da língua falada de crianças. À frequência média de palavras por frase há alternativas de cálculos, como, por exemplo, a frequência média de palavras por unidade-t e a frequência média de orações subordinadas por unidade-t, como se vê em Hunt (1965). Pode ainda ser uma medida a frequência média de palavras por oração, como proposto por Berman (2007). Também, Berman e Ravid (2009) referem que a complexidade sintática se pode obter alternativamente pela frequência média de orações num texto. McNamara *et al.* (2010), com base na plataforma Coh-Metrix, avaliam a complexidade sintática de um texto com base num conjunto de propriedades acumuláveis (e extraíveis automaticamente), sendo a mais relevante delas a média de palavras à esquerda do verbo principal de uma oração. Ainda, Nippold e Sun (2012) dividem a somatória de todas as orações (principais e subordinadas) pelo total de unidades-t num texto, medida a que chamam de densidade oracional.

Por serem indicadores amplamente referidos na literatura como fiáveis para a avaliação da extensão do desenvolvimento sintático (Biber *et al.* 2011), nesta tese, sigo os passos de Hunt (1965: 36) e Berman (2007: 351) e utilizo a frequência média de palavras por oração, a frequência média de palavras por unidade-t e a frequência média de orações subordinadas por unidade-t. Embora tais medidas não sejam livres de críticas, o seu emprego aqui também se justifica pela possibilidade de se compararem os resultados aqui encontrados com resultados de estudos anteriores, mesmo que baseados noutras línguas, de que se pode citar Verhoeven *et al.* (2001), Berman e Verhoeven (2002), Nippold *et al.* (2005) ou Beers e Nagy (2011).

Abaixo, encontra-se um fragmento de um texto argumentativo pertencente ao *corpus* sob estudo, segmentado em unidades-t (numeradas) e em orações subordinadas (ladeadas por parênteses retos), para ilustrar a aplicação das medidas acima referidas:

1. Eu acho [que hoje em dia é importante [ter computador mas não as redes sociais]].⁴¹
2. As redes sociais são boas [para conhecer pessoas].
3. Mas também começou a haver mais raptos e mortes.
4. Eu sou a favor e contra [porque há cada vez mais pessoas a [optarem por [sair da escola e de casa]]].
5. Mas é bom ter computador [porque os professores mandam trabalhos [para fazer no computador]].
6. E ficam melhores e apresentáveis.

Como está explícito no fragmento acima, trata-se de um texto com seis unidades-t, das quais duas são simples (3 e 6) – se (como aqui faço, por simplicidade) não assumirmos uma análise bioracional das estruturas com verbos auxiliares aspetuais, como as de (3) (*começou a haver...*) – e quatro são multioracionais (1, 2, 4 e 5), com catorze orações (seis independentes e oito subordinadas) e com 71 palavras. Extraí-se disto que há, em média,

⁴¹ É importante referir que, nesta tese, não estão a ser consideradas as estruturas de elipse, razão por que, na análise da unidade-t 1, a estrutura “mas não as redes sociais”, em que parece haver a elipse de *são importantes* (“mas não ~~são importantes~~ as redes sociais”), está incluída na oração anterior. Sem desmerecer a importância do recurso à elipse, que, por si só, legitimaria uma tese, em particular a partir de uma perspectiva desenvolvimental, a sua identificação (manual) numa quantidade considerável de textos, em paralelo com todos os outros itens de investigação que aqui se utilizam, implica um esforço que ultrapassa em muito as possibilidades definidas nesta tese, para além da subjectividade envolvida neste tipo de análise.

5,07 palavras por oração e 11,83 palavras por unidade-t e que a frequência média de orações subordinadas por unidade-t é de 1,33.

É importante ressaltar que o procedimento com base na extensão não deve ser entendido como um mero contar de unidades. Como afirmam Beers e Nagy (2009: 186), adotar este tipo de procedimento para a análise de um texto ajuda a revelar o aspecto qualitativo da complexidade sintático-lexical. Estes autores fundamentam-se na lógica plana de que uma proporção elevada de palavras por oração, e por analogia por unidade-t, implica um maior encapsulamento de informações. Asseguram ainda que estas frequências refletem a interface entre estruturação sintática e literacia linguística, representada no emprego de nominalizações, adjetivos atributivos ou sintagmas preposicionais. A interface ‘léxico-sintaxe’ explicita-se bem nas palavras de Ravid (2005: 343):

Longer clauses contain more words in general, and more content words in particular, and so they are lexically denser than shorter clauses. Besides, more words per clause translate into larger phrases (NPs, VPs, and PPs) in the clause, and this in turn is indicative of a more complex and diverse clause-internal syntactic architecture.

5.3. Variação

O procedimento com base na variação, ou qualitativo, faz corresponder a complexidade sintática à subordinação. É este o sentido mais tradicional do termo “complexidade” nos estudos linguísticos, em que a complexidade representa o encaixamento de uma unidade oracional dentro de outra unidade oracional, estabelecendo-se uma hierarquia entre essas unidades (Schleppegrell 1992: 117). Quer isto dizer objetivamente que quanto mais estruturas de subordinação um texto possui mais complexo ele é. É nisto que crêem vários autores. Kalmar (1985, cf. Schleppegrell 2004: 14) é contundente ao afirmar que a manifestação da dependência oracional indica níveis de evolução no uso de uma língua. Beers e Nagy (2011: 184) consideram que certos tipos de oração dependente permitem aos escritores expressarem ideias mais elaboradas. Biber (1992: 139), um pouco mais cauteloso, sugere que a presença de dependência oracional num texto representa o aumento da complexidade estrutural apenas, estando a complexidade global de um texto associada a um conjunto mais amplo de indicadores, que inclui, por exemplo, a especificidade lexical.

A importância atribuída à dependência no processo de desenvolvimento da língua escolar atesta-se ainda na seguinte afirmação de Loban (1963, cf. Schleppegrell 2004: 14): “Both logical analysis and previous studies of language designate subordination as a more mature and difficult form of language expression than simple parallel statements connected by and or but”. Das palavras de Loban, conclui-se que crianças que empregam mais orações subordinadas nos seus textos são mais bem-sucedidas do que as que não o fazem. A mesma ideia é corroborada por Givón (1991: 347). Beaman (1984: 45) defende que textos mais simples contêm mais orações coordenadas do que subordinadas e, contrariamente, textos complexos contêm mais orações subordinadas do que coordenadas.

Para avaliar a variação da complexidade sintática, muitas são as propostas de técnicas de medição, originárias dos estudos de desenvolvimento ora em L1, ora em L2. A mais tradicional, que se vê originalmente em La Brant (1933), consiste na razão entre o número de orações subordinadas e o número total de orações (subordinadas e independentes), sendo a frase a unidade considerada como o escopo da ocorrência de orações. Tal técnica tornou-se conhecida como índice de subordinação (*subordination index*). Apesar das muitas tentativas de reformulação da técnica original de La Brant, até mesmo com a incorporação de princípios da gramática transformacional, como se vê em Fodor, Bever e Garrett (1974), que propõem medir a complexidade das frases pelo número de transformações (proposta conhecida como Teoria da Complexidade Derivacional), ecoados em Loban (1963) ou em O'Donnell *et al.* (1967), até hoje o rácio de subordinação continua a ser uma forma de medição bastante recorrente nos estudos sobre o desenvolvimento sintático. Para esta tese, aplica-se a medida original, que é a razão de orações subordinadas pelo total de orações, mas fazem-se dois ajustes. O primeiro deles, seguindo Hunt (1965), é utilizar a unidade-t, e não a frase, como escopo. O segundo é considerar a especialidade da oração subordinada quanto à função, se completiva, adverbial, relativa ou gradativa, o que atribui à medida um caráter qualitativo e justifica a separação desta medida da sua contraparte genérica, que é a frequência média de orações por unidade-t.

Para a identificação e anotação dos tipos de orações subordinadas, que estão mais bem descritas e exemplificadas no Anexo 9.7 (Tipos e subtipos de oração subordinada), seguem-se as análises de Móia (1996), Duarte (2003), Brito e Duarte (2003), Brito (2003), Matos (2003), Peres e Mascarenhas (2006), Perini (2010), Marques (2013), Mendes (2013), Barbosa (2013), Barbosa e Raposo (2013), Lobo (2013) e Veloso (2013), entre outros. São

também referências Quirk *et al.* (1985), Greenbaum (1996), Huddleston e Pullum (2002, 2006) e Biber *et al.* (2007). Tendo em conta estes estudos, consideram-se quatro tipos básicos de orações subordinadas: completivas, adverbiais, relativas e gradativas: i) as orações completivas funcionam como argumentos selecionados por um núcleo, que pode ser verbal, nominal ou adjetival; ii) as orações adverbiais funcionam como adjuntos – não-argumentos – de outra oração, pelo que não são selecionadas por qualquer elemento da oração a que se subordinam, funcionando como adjuntos de um predicado (ou sintagma ou frase); iii) as orações relativas funcionam ou como modificadores na estrutura do sintagma nominal ou como comentários (i.e. apostos) a um sintagma nominal ou frase, a que se ligam; por fim, iv) as orações gradativas expressam informação associada a uma ou mais escalas relevantes no significado de nomes, verbos, advérbios ou adjetivos que são passíveis de gradação.

Retome-se o fragmento de texto utilizado acima, desta vez, para exemplificar a frequência e os rácios de subordinação:

1. Eu acho [que hoje em dia é importante [ter computador mas não as redes sociais]].
2. As redes sociais são boas [para conhecer pessoas].
3. Mas também começou a haver mais raptos e mortes.
4. Eu sou a favor e contra [porque há cada vez mais pessoas a [optarem por [sair da escola e de casa]]].
5. Mas é bom ter computador [porque os professores mandam trabalhos [para fazer no computador]].
6. E ficam melhores e apresentáveis.

Como já referido, neste fragmento há um total de seis unidades-t e de catorze orações, das quais seis são independentes e oito subordinadas. Quanto ao tipo de subordinação, têm-se: quatro completivas ([que hoje em dia é importante ter computador mas não as redes sociais], [ter computador mas não as redes sociais], [para conhecer pessoas] e [sair da escola e de casa]), dois modificadores infinitivos, com função adjetiva afim da de uma oração relativa (e que por facilidade incluirei no mesmo grupo) ([para) fazer no computador] e [(a) optarem por sair da escola e de casa]) e duas adverbiais ([porque há cada vez mais pessoas a optarem por sair da escola e de casa] e [porque os professores mandam trabalhos]), de que resulta um rácio de 0,57 subordinadas por oração. Consideradas as subclasses de orações subordinadas pelo total de orações subordinadas, as taxas de rácio indicam que há 0,50 completivas, 0,25 relativas (ou afins) e 0,25 adverbiais para cada oração subordinada. Quer isto dizer, noutra forma de expressão, que, do total de orações, 57% são subordinadas e que, do total de orações subordinadas, 50% das orações são completivas, 25% são relativas (ou afins) e 25% são adverbiais.

Em síntese, as medidas de composição da complexidade sintática utilizadas nesta tese são a frequência média de palavras por oração, a frequência de palavras por unidade-t e a frequência média de orações subordinadas por unidade-t – para avaliar a extensão – e o rácio de orações subordinadas (completivas, adverbiais, relativas e gradativas) pelo total de orações – para avaliar a variação.

5.4. Procedimento metodológicos específicos

Todo o procedimento de segmentação e anotação das orações e unidades-t realizou-se no programa informático UAMCorpusTool (O'Donnell 2007), doravante UCT. A opção por esta plataforma justifica-se num conjunto de razões de carácter operacional. Em primeiro lugar, o UCT é uma plataforma de interface funcional, que permite a segmentação manual de unidades em qualquer extensão e que também permite o manuseio, com relativa facilidade, dos segmentos identificados e respetivas etiquetas de anotação. Em segundo lugar, o UCT permite a criação e a edição de esquemas de anotação em forma de sistema, com múltiplas entradas de classificação em vários níveis, podendo estes esquemas ser especificados pelas necessidades particulares de uma dada investigação (Figura 1). Por fim, esta plataforma permite a consulta (*query*) não só pelas etiquetas de anotação, como também por *wildcards*. Apresenta-se na figura que se segue a classificação de orações considerada:

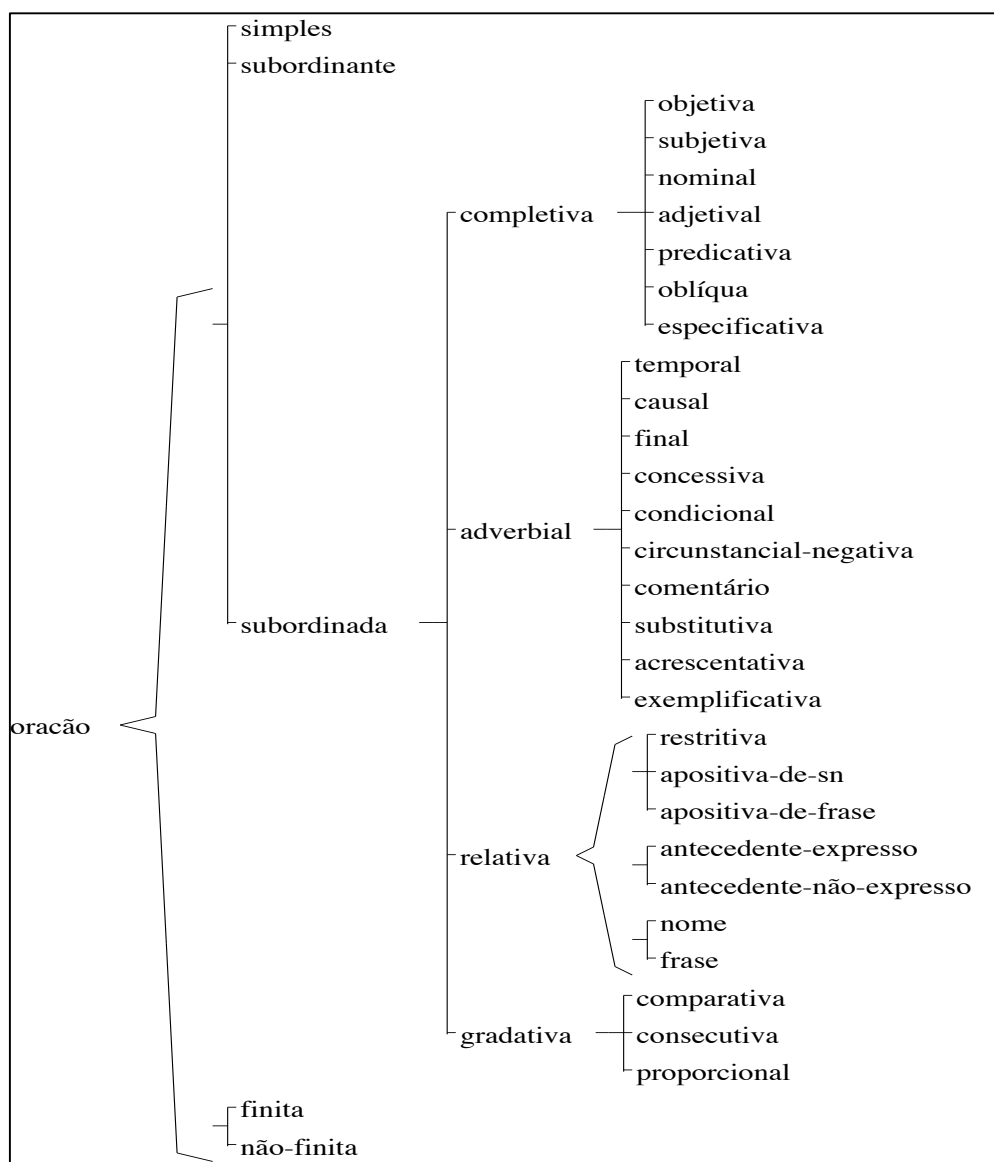


Figura 1: Sistema desenhado na plataforma UCT para a anotação dos tipos e subtipos de orações subordinadas conforme descritos no Anexo 9.7.

Todos os textos foram segmentados em unidades-t e em orações, já que estas se constituem como unidades fundamentais a todos os aspetos da complexidade sintática que serão aqui estudados. Nesta tese, como já referido, oração é “uma sequência de palavras gramatical, que tem um conteúdo proposicional e como elemento nuclear um verbo” (Raposo 2013: 314). Uma apresentação mais detalhada dos tipos de orações subordinadas consideradas nesta tese faz-se no Anexo 9.7.

As orientações dos autores acima mencionados conformam-se como um parâmetro da variedade culta (adulta e escolarizada) da língua. É ainda importante referir que a classificação baseada nestes autores não se pretendeu exaustiva, pelo que não foram incluídos todos os tipos de subordinação que eles descrevem. A classificação aqui utilizada tenta dar conta dos casos de realização identificados no *corpus*. Acresce a isto o facto de, a par da complexidade sintática, haver outros indicadores de complexidade linguística que aqui se adotam, o que impõe alguns limites a esta investigação para que ela seja exequível, já que o trabalho depende, em grande parte, de anotação manual, isto é, não-automática. Esta também é a razão por que as orações coordenadas e suplementares (globalmente tratadas como simples) se consideram apenas para fins de distinção das subordinadas, sem classificações pormenorizadas.

No UCT, para os efeitos de anotação pretendidos, os segmentos oracionais foram divididos em dois tipos: o segmento subordinante, que contém uma oração matriz; e o segmento subordinado, que contém uma oração dependente de outra. No caso de a oração ser bifuncional, em que é, por exemplo, subordinada de uma oração e subordinante de outra, prevaleceu a anotação como 'subordinada'. Se é coordenada com uma oração e subordinante de outra, prevaleceu a anotação como 'subordinante'. Justifica-se isto por não ser possível a atribuição de duas ou mais etiquetas a um único segmento no UCT. As orações completivas, relativas e gradativas, que são nomeadas de segmentos subordinados, estão incluídas no segmento da oração em que está localizado o constituinte local a que estão associadas (e.g., a estrutura nominal que modificam, no caso das orações relativas restritivas; o predicador que as seleciona, no caso das completivas), que constitui o segmento subordinante. As orações adverbiais também constituem segmentos subordinados, mas estão segmentadas separadamente (não-encaixadas), já que a dependência é "mais ténue do que nas outras estruturas de subordinação" (Lobo 2013: 1989). Na terminologia de Peres e Mascarenhas (2006), estas são estruturas de subordinação livre, ao passo que aquelas são de subordinação presa.

A unidade-t é, lembrando Hunt (1965: 21), a unidade constituída por uma oração principal e, quando for o caso, da(s) oração(ões) que a ela se subordina(m). É importante salientar que, seguindo Hunt (1965) e autores posteriores, como Loban (1976) ou Biber *et al.* (2007), cada oração integrada em compostos coordenados constitui uma unidade-t *per se*. Estes autores justificam-no, entre outras razões, com o facto de a oração iniciada por coordenadores habitualmente conformar-se como frase autónoma, em particular na escrita, a

despeito de não ser esta a posição assumida nas gramáticas tradicionais. Hunt (1965: 25) distingue ainda as unidades-t em dois tipos: simples e multioracionais. A unidade-t simples realiza-se com apenas uma oração, que é, obviamente, independente, ou seja, é o segmento que contém uma oração não-subordinada e não-subordinante (simples, coordenada, suplementar – ou âncora – ou justaposta). As unidades-t, no UCT, foram anotadas em conformidade com esta distinção.

A Tabela 4, abaixo, dispõe dos valores absolutos obtidos após o procedimento de segmentação e anotação, que foi feito de forma inteiramente manual, em virtude de não haver para o português ferramentas computacionais capazes de segmentar e anotar textos automaticamente nos níveis da oração e da unidade-t de modo fiável. Os valores correspondem às ocorrências de unidades-t na totalidade (*unt*), de unidades-t simples (*sim*), de unidades-t multioracionais (*mul*), de orações na totalidade (*orc*) e de orações subordinadas (*sub*). Para tornar abreviada a leitura dos dados, recupera-se na tabela o número total de palavras (*tok*) identificado no *corpus*:

Registo	ano	unt	sim	mul	orc	sub	tok
Narrativo	5.º	560	387	173	819	259	4 657
	7.º	962	632	330	1 468	506	8 305
	10.º	928	506	422	1 700	772	10 381
Argumentativo	5.º	328	138	190	698	370	4 016
	7.º	519	228	291	1 104	585	7 049
	10.º	639	285	354	1 346	707	9 277
Total	-	3 936	2 176	1 760	7 135	3 199	43 685

Tabela 4: Valores absolutos das ocorrências de unidades-t (*unt*), unidades-t simples (*sim*), unidades-t multioracionais (*mul*), orações (*orc*), orações subordinadas (*sub*) e palavras (*tok*) nos registos narrativo e argumentativo.

Em virtude de os dados obtidos a partir da segmentação e anotação não obedecerem a uma distribuição normal, nem serem passíveis de normalização, pela grande quantidade de valores aberrantes, tornou-se necessária a utilização de testes não paramétricos. Assim, para a identificação da magnitude de correlação entre as variáveis, aplicou-se o coeficiente de correlação de Spearman, sendo corroborado, para detetar a existência de variância significativa na amostra, pelo teste de Kruskal-Wallis (análogo ao teste F da ANOVA a um fator),

de que também se extraem a força do tamanho do efeito e os valores de probabilidade de diferenças entre os grupos.

5.5. Estudo de correlação

Por serem dois os aspetos da complexidade sintática que estão a ser aqui avaliados (a extensão e a variação), dividiu-se a apresentação dos resultados dos testes estatísticos, por uma questão de clareza e objetividade, em dois momentos distintos. Primeiramente são descritos os resultados quanto às medidas que representam a extensão da complexidade sintática. De seguida, é a vez dos resultados relativos às medidas de variação. Após a apresentação dos resultados de ambos os momentos, têm lugar a discussão e as considerações sobre os resultados na totalidade.

Na Tabela 5, abaixo, apresentam-se, diferenciados por registo (narrativo e argumentativo) e por ano escolar (5.º, 7.º e 10.º), os números médios de ocorrência de orações (*Orc*), de orações subordinadas (*Sub*), de unidades-t (*Unt*) e de palavras (*Tok*) – estes repetidos do capítulo anterior para facilitar a leitura dos resultados –, com os respetivos desvios-padrão, identificados nos textos sob estudo. São estas as unidades centrais para a medição da extensão da complexidade sintática.

Registo	Ano	N	Tok (DP)		Orc (DP)		Sub (DP)		Unt (DP)	
Narrativo	5.º	26	179,12	45,14	31,88	11,70	9,88	5,02	21,54	8,94
	7.º	46	180,54	37,32	32,26	8,71	10,96	5,55	20,91	5,18
	10.º	50	207,62	48,40	34,68	10,25	15,36	6,34	18,56	5,55
Argumentativo	5.º	26	154,46	39,26	27,27	8,62	14,19	5,91	12,62	4,39
	7.º	46	153,24	38,26	23,98	5,75	12,89	4,47	11,28	4,35
	10.º	50	185,54	32,11	34,68	10,25	13,98	4,94	12,78	3,42

Tabela 5: Valores médios de ocorrência, por texto, de palavras (*Tok*), de orações (*Orc*), de orações subordinadas (*Sub*) e de unidades-t (*Unt*) identificados nos registos narrativo e argumentativo.

No registo narrativo, a quantidade média de palavras por texto é crescente de ano a ano. Também o são o número médio de ocorrências de orações (simples, subordinante e subordinada) e o número médio de ocorrências de orações subordinadas. Num movimento contrário, o número médio de unidades-t é decrescente do 5.º ao 7.º, e deste ao 10.º. Nos textos do registo argumentativo, o 7.º ano escolar é o ponto que anula a linearidade das

ocorrências das quatro unidades, já que o número médio de ocorrências é sempre mais baixo neste ano do que nos outros, isto é, do 5.º ao 7.º, há um decréscimo nos valores médios, mas deste ao 10.º ano vê-se uma elevação.

Os valores acima descritos servem de base ao cálculo das três medidas de avaliação da extensão da complexidade sintática escolhidas para este estudo, que são a frequência média de palavras por oração (*Tok/Orc*), a frequência média de palavras por unidade-t (*Tok/Unt*) e a frequência média de orações subordinadas por unidade-t (*Sub/Unt*), sendo a primeira originada em Berman (2007) e as duas últimas em Hunt (1965, 1970). O Gráfico 5 e o Gráfico 6, abaixo, apresentam o comportamento pareado destas frequências ao longo dos anos escolares no registo narrativo e no registo argumentativo, respetivamente:

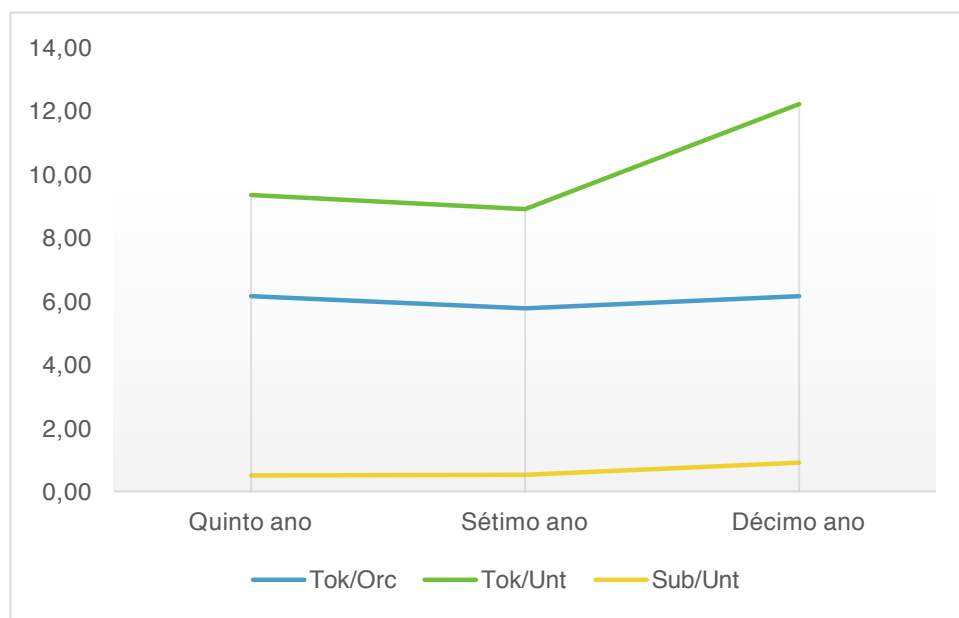


Gráfico 5: Frequência média de palavras por oração (*Tok/Orc*), de palavras por unidade-t (*Tok/Unt*) e de orações subordinadas por unidade-t (*Sub/Unt*) no registo narrativo ao longo dos anos escolares.

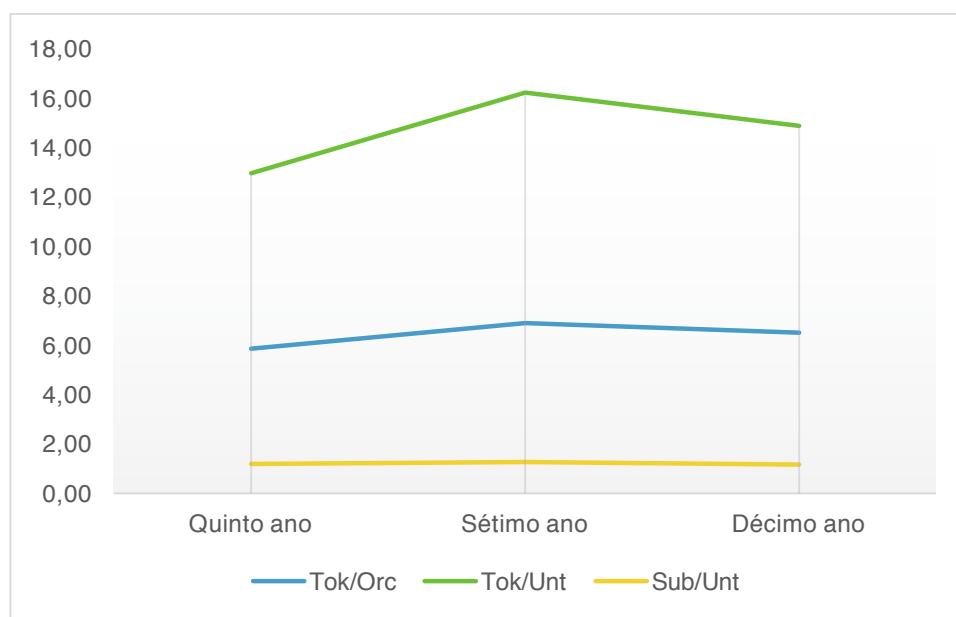


Gráfico 6: Frequência média de palavras por oração (*Tok/Orc*), de palavras por unidade-t (*Tok/Unt*) e de orações subordinadas por unidade-t (*Sub/Unt*) no registo argumentativo ao longo dos anos escolares.

Quanto à frequência média de palavras por oração, percebe-se que os valores médios, no registo narrativo (Gráfico 5), diminuem do 5.º para o 7.º ano, mas aumentam deste ano para o 10.º ano. No 5.º ano, a frequência média de palavras por oração é de 6,15 (2,05⁴²); no 7.º ano, o valor médio é de 5,77 (0,94); no 10.º ano, o valor médio é de 6,16 (1,00). Quanto ao registo argumentativo (Gráfico 6), o movimento é precisamente o oposto, havendo uma ligeira subida do 5.º ao 7.º ano. Deste ano para o 10.º, pelo contrário, ocorre uma diminuição, também ligeira. No 5.º ano, a frequência média de palavras por oração é 5,84 (1,10); no 7.º ano, o valor médio é de 6,89 (1,29); no 10.º ano, o valor médio é de 6,54 (1,20).

A frequência média de palavras por unidade-t reflete um cenário bastante diferenciado da frequência anterior. Quanto ao registo narrativo, o Gráfico 5 sugere que há alguma diminuição nos valores do 5.º para o 7.º ano; porém, verifica-se do 7.º para o 10.º ano uma subida bastante acentuada no uso. A frequência média de palavras por unidade-t, no 5.º ano, é de 9,09 (2,98); no 7.º ano, a frequência média é de 9,09 (2,05); no 10.º ano, o valor médio é de 12,12 (5,96). No registo argumentativo (Gráfico 6), a frequência média de palavras por

42 Desvio-padrão.

unidade-t também se mostra diferenciada da frequência de palavras por orações no mesmo registo. Do 5.º ao 7.º ano, há um aumento acentuado na frequência de palavras por unidade-t, seguido de uma diminuição também acentuada deste para o 10.º ano. No 5.º ano, a frequência média é de 12,94 (3,08); no 7.º ano, é de 16,21 (6,29); e, no 10.º ano, é de 14,86 (5,27).

O Gráfico 5 mostra, quanto à frequência média de orações subordinadas por unidade-t no registo narrativo, que há um aumento gradativo nos valores do 5.º ao 10.º ano, mesmo que aparentemente ténue. No 5.º ano, a frequência média de orações subordinadas por unidade-t é de 0,51 (0,27); no 7.º ano, o valor médio passa a 0,54 (0,28); no 10.º ano, o valor médio é de 0,92 (0,61). No registo argumentativo, a partir do Gráfico 6, infere-se que os valores mantêm-se quase inalterados de ano a ano, havendo uma subida pouco expressiva dos valores no 7.º ano, quando comparados com os dos outros anos escolares. No 5.º ano, o valor médio de orações subordinadas por unidade-t é de 1,18 (0,46); no 7.º ano, a frequência sobe para 1,27 (0,68); e, no 10.º ano, cai para 1,16 (0,45).

Os testes de correlação aplicados às três medidas corroboram a leitura que se pode fazer dos valores médios apresentados no Gráfico 5 e no Gráfico 6. Para a relação entre a frequência de palavras por oração e a progressão escolar não se identificou correlação matematicamente significativa nem no registo narrativo ($r=0,176^{43}$; $p=0,052^{44}$), nem no registo argumentativo ($r=0,154$; $p=0,090^{45}$). Mesmo sem correlação, é importante referir que, pelo teste de variância, foram notadas diferenças significativas, nas médias do registo narrativo, entre o 7.º e o 10.º ano ($p=0,029$), não havendo diferenças nem entre o 5.º e o 7.º ano ($p=0,930$), nem entre o 5.º e o 10.º ano ($p=0,156$). Quanto às médias dos textos argumentativos, as diferenças localizam-se entre o 5.º e o 7.º ano ($p=0,000$) e entre o 5.º e o 10.º ano ($p=0,004$), não havendo diferenças entre o 7.º e o 10.º ano ($p=0,080$).

⁴³ Para a interpretação da magnitude das correlações em Spearman, utiliza-se a classificação dos coeficientes de correlação proposta por Hulley *et al.* (2003): $< 0,4$ (correlação fraca), $\geq 0,4$ a $\leq 0,5$ (correlação moderada) e $> 0,5$ (correlação forte).

⁴⁴ Diferença significativa no nível $<0,01$.

⁴⁵ Correlação significativa no nível $<0,01$ (2 extremidades).

Para a frequência de palavras por unidade-t no registo narrativo, identificou-se uma correlação positiva moderada ($r=0,433$; $p=0,000$). Quanto a isto, na análise da variância pelo teste de Kruskal-Wallis, foram detetadas diferenças significativas entre as médias dos três grupos tomados conjuntamente ($H(2)=26,968$; $p=0,000^{46}$), com os seguintes valores de *rank* – no 5.º, 47,46; no 7.º, 47,75; no 10.º, 81,45 – e com um tamanho do efeito de 0,22. No entanto, confrontados os valores de *rank* de cada grupo, não se detetou diferença entre o 5.º e o 7.º ano ($p=0,699$), diferença que se manifesta apenas entre o 5.º e o 10.º ano ($p=0,000$) e entre o 7.º e o 10.º ano ($p=0,000$). Quanto à frequência de palavras por unidade-t no registo argumentativo e à progressão escolar, não foi identificada nenhuma correlação significativa ($r=0,077$; $p=0,402$).

Por fim, os testes estatísticos apontam para uma correlação positiva moderada entre a frequência média de orações subordinadas por unidade-t nos textos do registo narrativo e a progressão escolar ($r=0,446$; $p=0,000$), sendo tal correlação confirmada pela existência de diferenças significativas entre as médias dos anos escolares ($H(2)=26,918$; $p=0,000$), com os seguintes valores de *rank* – no 5.º ano, 45,35; no 7.º ano, 49,04; no 10.º ano, 81,36 – e com um tamanho do efeito de 0,22. É de referir, no entanto, que entre o 5.º e o 7.º ano, não foram identificadas diferenças significativas ($p=0,606$), que se verificam apenas entre o 7.º e o 10.º ano ($p=0,000$) e entre o 5.º e o 10.º ano ($p=0,000$). Quanto aos textos argumentativos, não foi identificada nenhuma correlação entre a frequência média de orações subordinadas por unidade-t e a progressão nos anos escolares ($r=-0,009$; $p=0,920$).

Apresentados os resultados dos testes estatísticos aplicados às medidas de extensão da complexidade sintática, seguem-se, agora, os resultados relativos à variação. Para tanto, estão dispostos na Tabela 6, abaixo, os valores médios de ocorrências, por texto, das orações subordinadas pelos grandes quatro tipos: completivas (*com*), adverbiais (*adv*), relativas (*rel*) e gradativas (*gra*). Todos estes valores são basilares à compreensão dos resultados dos rácios de subordinação.

Registo	Ano	N	Com (DP)		Adv (DP)		Rel (DP)		Gra (DP)	
Narrativo	5.º	26	4,23	3,66	4,15	2,40	1,08	1,08	0,08	0,27

⁴⁶ Diferença significativa no nível $<0,05$.

Argumentativo	7.º	46	4,26	2,96	5,26	2,51	1,91	1,91	0,07	0,25
	10.º	50	6,00	3,23	4,60	2,81	4,22	4,22	0,18	0,44
	5.º	26	5,15	2,96	4,69	2,21	4,54	4,54	0,15	0,46
	7.º	46	4,80	2,79	4,59	2,39	4,15	4,15	0,07	0,25
	10.º	50	5,24	3,56	3,96	2,83	3,92	3,92	0,02	0,14

Tabela 6: Valores médios de ocorrência, por texto, de orações completivas (*Com*), adverbiais (*Adv*), relativas (*Rel*) e gradativas (*Gra*) identificados nos registos narrativo e argumentativo.

No registo narrativo, como se depreende dos valores apresentados na Tabela 6, o emprego de orações completivas mostra-se crescente do 5.º ao 7.º ano, e deste ao 10.º ano de escolaridade. As orações adverbiais não seguem o padrão do tipo anterior, já que, embora haja um aumento no seu emprego do 5.º ano para o 7.º, há uma diminuição deste para o 10.º ano. As orações relativas apresentam um uso sempre crescente ao longo dos anos de escolaridade. Por fim, as orações gradativas, que apresentam valores de uso muito pouco expressivos em todos os anos, têm valores de ocorrência próximos entre o 5.º e o 7.º ano, com algum aumento no 10.º ano. Diferentemente do registo narrativo, no registo argumentativo, o emprego de orações completivas é mais pequeno no 7.º ano, em comparação com os outros anos. Os valores médios das orações adverbiais, relativas e gradativas descrevem-se num processo de diminuição linear do ano inicial até ao ano final.

Os valores do rácio de subordinação por tipo (*Com/Sub*, *Adv/Sub*, *Rel/Sub* e *Gra/Sub*) identificados nos textos apresentam-se no Gráfico 7 e no Gráfico 8 referentes respetivamente, ao registo narrativo e ao argumentativo. A esta apresentação, seguem-se os resultados da aplicação dos testes estatísticos.

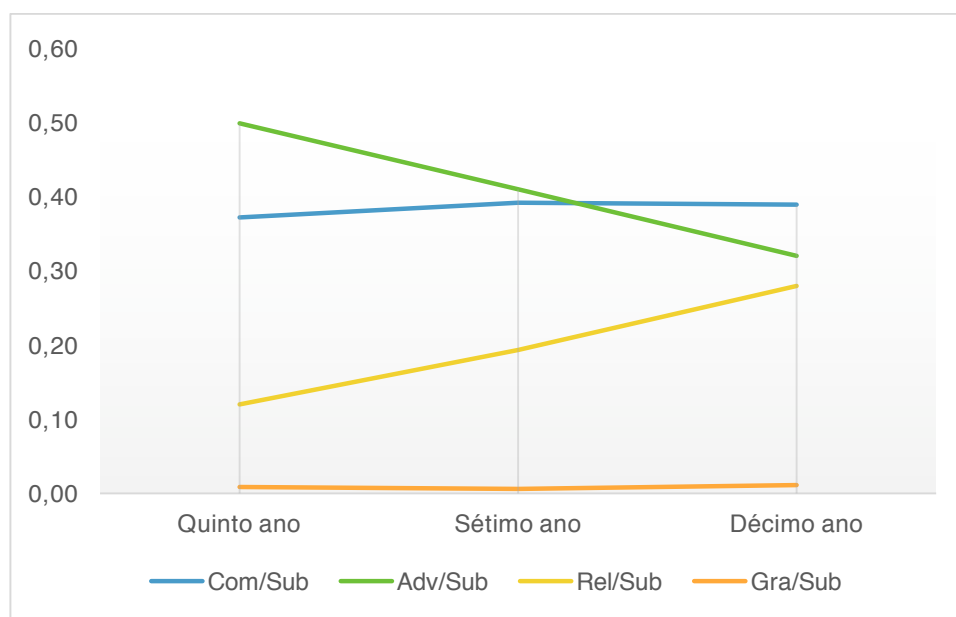


Gráfico 7: Valores médios do rácio de orações completivas (*Com*), orações adverbiais (*Adv*), orações relativas (*Rel*) e orações gradativas (*Gra*) pelo total de orações subordinadas (*Sub*) identificados nos textos do registo narrativo.

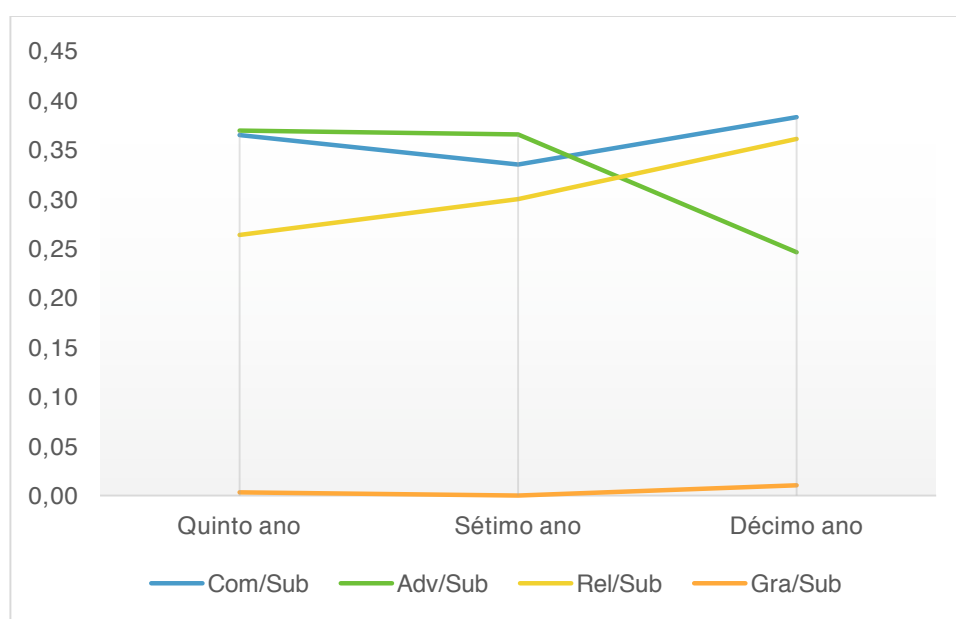


Gráfico 8: Valores médios do rácio de orações completivas (*Com*), orações adverbiais (*Adv*), orações relativas (*Rel*) e orações gradativas (*Gra*) pelo total de orações subordinadas (*Sub*) identificados nos textos do registo argumentativo.

O Gráfico 7 ilustra o movimento dos rácios de subordinação nos textos do registo narrativo, sugerindo que há bastante heterogeneidade no comportamento das variáveis. Os rácios de utilização das orações completivas mantêm-se semelhantes entre os anos, com um

aparente aumento do 5.º para o 7.º ano e sem muitas diferenças deste ano para o 10.º. No 5.º ano, o rácio de completivas pelo total de orações subordinadas é de 0,37 (0,22); no 7.º ano, este valor passa a 0,39 (0,20); e, no 10.º ano, o rácio é de 0,39 (0,17). O uso de adverbiais decai expressiva e constantemente do 5.º para o 10.º. No 5.º ano, o rácio de adverbiais pelo total de orações subordinadas é de 0,50 (0,23); no 7.º ano, é de 0,40 (0,19); e, no 10.º ano, o rácio é de 0,32 (0,16). As orações relativas, como o Gráfico 7 sugere, são o único tipo que se mostra de uso linearmente crescente de ano a ano. No 5.º ano, o rácio de relativas pelo total de orações subordinadas é de 0,12 (0,13); no 7.º ano, este valor passa a 0,19 (0,13); e, no 10.º ano, o rácio é de 0,28 (0,16). Por fim, o uso de orações gradativas não exhibe diferenças notáveis ao longo dos anos. No 5.º ano, o rácio de gradativas pelo total de orações subordinadas é de 0,01 (0,03); no 7.º ano, este valor continua a 0,01 (0,03); e, no 10.º ano, o rácio também é de 0,01 (0,03).

No registo argumentativo, conforme se depreende do Gráfico 8, é possível notar que a heterogeneidade quanto ao uso das orações subordinadas completivas, adverbiais e relativas é mais evidente ainda do que no registo narrativo, sendo a exceção as orações gradativas, que mantêm o mesmo comportamento ao longo dos anos. Como se vê, há uma diminuição no uso das orações completivas do 5.º para o 7.º ano, seguido de um aumento deste ano para o 10.º. No 5.º ano, o rácio de orações completivas pelo total de orações subordinadas é de 0,36 (0,15); no 7.º ano, este rácio baixa para 0,33 (0,19); e, no 10.º, eleva-se para 0,38 (0,16). O uso de orações adverbiais, como sugerido pelo Gráfico 8, diminui de modo ténue do 5.º para o 7.º ano, sendo mais expressiva a diminuição entre este e o 10.º ano da escolaridade. O rácio de orações adverbiais, no 5.º ano, é de 0,37 (0,12); no 7.º, baixa para 0,36 (0,15); e, no 10.º ano, baixa mais fortemente para 0,25 (0,12). Quanto às relativas, tem-se um cenário semelhante ao que acontece nos textos do registo narrativo, havendo um crescimento no uso de ano a ano. No 5.º ano, o rácio de orações relativas é de 0,26 (0,11); no 7.º, a taxa sobe para 0,30 (0,15); e, no 10.º ano, continua a elevar-se, atingindo a taxa de 0,36 (0,17). As orações gradativas, em virtude de serem ocorrências numéricas de valores muito baixo, ficam-se, no 5.º ano, pela taxa de 0,00 (0,01); no 7.º, não houve ocorrências; no 10.º ano, tem-se uma taxa de 0,01 (0,03).

Os testes estatísticos aplicados sobre os valores de rácio corroboram a leitura que se faz dos gráficos acima. No registo narrativo, identificou-se que não há correlação entre a proporção de orações completivas e a progressão escolar ($r=0,004$; $p=0,964$). Em relação à

proporção de orações subordinadas adverbiais, há uma correlação negativa fraca com a progressão escolar, ou seja, quanto mais o aluno avança nos anos escolares, menos utiliza este tipo de subordinada ($r=-0,325$; $p=0,000$). Confirma-o a existência de diferenças significativas nas médias reveladas pela análise de variância no teste de Kruskal-Wallis ($H(2)=12,759$; $p=0,002$), com os seguintes valores de *rank* – no 5.º ano, 78,56; no 7.º, 65,34; no 10.º, 49,10 – e com um tamanho do efeito de 0,11. Os valores de *rank* de cada grupo apontam para a inexistência de diferenças entre o 5.º e o 7.º ano ($p=0,121$), havendo apenas entre o 5.º e o 10.º ano ($p=0,001$) e entre o 7.º e o 10.º ano ($p=0,024$). Quanto à ocorrência de orações relativas, os testes revelam uma correlação positiva moderada ($r=0,425$; $p=0,000$), corroborada por diferenças entre as médias dos grupos estudados ($H(2)=22,007$; $p=0,000$), com os seguintes valores de *rank* – no 5.º, 38,31; no 7.º, 57,34; no 10.º, 77,39 –, com um tamanho do efeito amostral de 0,18. Pelos valores de *rank*, têm-se diferenças estatisticamente significativas entre todos os anos: entre o 5.º e o 7.º ano ($p=0,014$), entre o 7.º e o 10.º ($p=0,003$) e entre o 5.º e o 10.º ano ($p=0,000$). Para as orações gradativas, não foi identificada correlação com a progressão escolar ($r=0,116$; $p=0,203$).

As taxas de rácio identificadas no registo argumentativo constroem um cenário estatístico semelhante ao que se viu no registo narrativo. A relação entre a proporção de orações completivas e a progressão escolar não se revela significativa, pelo que não há correlação entre essas variáveis ($r=0,092$; $p=0,313$). A comparação dos valores do rácio de orações adverbiais pelo total de orações subordinadas e a progressão escolar, por sua vez, resultam numa correlação negativa fraca ($r=-0,390$; $p=0,000$). Na análise da variância, deteta-se que há diferenças entre os valores realizados pelos grupos ($H(2)=20,449$; $p=0,000$), com os seguintes valores de *rank* – no 5.º, 75,81; no 7.º, 72,20; no 10.º, 44,22 –, com um tamanho do efeito de 0,17. A comparação entre os grupos, no entanto, revela que não há diferenças entre o 5.º e o 7.º ano ($p=0,681$), contrariamente ao que se verifica entre o 7.º e o 10.º ano ($p=0,000$) e entre o 5.º e o 10.º ano ($p=0,000$). Do mesmo modo que acontece no registo narrativo, no argumentativo, a utilização das orações relativas em relação ao total de orações subordinadas vê-se numa correlação positiva, mas desta vez fraca, com a progressão escolar ($r=0,222$; $p=0,014$), o que se confirma com a existência de diferenças nas médias dos grupos estudados ($H(2)=5,999$; $p=0,050$), com os seguintes valores de *rank* – no 5.º, 49,58; no 7.º, 59,11; no 10.º, 69,90 –, com um tamanho do efeito de 0,05. Não foram detetadas diferenças nem entre o 5.º e o 7.º ano ($p=0,234$), nem entre o 7.º e o 10.º ($p=0,120$),

sendo a diferença localizada apenas entre o 5.º e o 10.º ano ($p=0,022$). A proporção de orações gradativas pelo total de subordinadas mantém uma correlação positiva fraca com a progressão escolar ($r=0,181$; $p=0,046$), com diferenças estatísticas entre os valores dos grupos na totalidade ($H(2)=6,561$; $p=0,038$), com os seguintes valores de *rank* – no 5.º, 60,33; no 7.º, 58,00; no 10.º, 65,33 –, com um tamanho do efeito de 0,05. Não há diferenças significativas nem entre o 5.º e o 7.º ano ($p=0,183$), nem entre o 5.º e o 10.º ano ($p=0,243$), havendo-as apenas entre o 7.º e o 10.º ano ($p=0,016$).

5.6. Discussão

Os resultados dos testes estatísticos indicam, de um modo geral, que a extensão das unidades linguísticas selecionadas – orações e unidades-t – não se conforma a um mesmo processo uniforme e linear de desenvolvimento ao longo dos anos escolares. Por um lado, a extensão da oração, medida pela frequência média de palavras, não mantém correlação com a progressão escolar em nenhum dos dois registos avaliados e, por outro, a unidade-t, quer aferida pela frequência média de palavras, quer pela frequência média de orações subordinadas, só tem representação significativa nos textos do registo narrativo, não sendo detetado o mesmo para os textos argumentativos. Estes resultados confirmam parcialmente a primeira hipótese de trabalho, segundo a qual a extensão sintática se correlaciona positivamente com a progressão nos anos escolares em ambos os registos. No entanto, confirmam plenamente a segunda hipótese de trabalho, segundo a qual os textos narrativos apresentam mudanças na sua caracterização sintática de modo mais expressivo do que as identificadas nos textos de registo argumentativo.

Com base apenas na extensão da oração por palavras nos dois registos apreciados, os resultados aqui encontrados contrariam uma convicção bastante difundida na literatura (e corroborada também pelos resultados de alguns estudos) sobre o desenvolvimento linguístico em idade escolar. Esta convicção pode-se sintetizar no que afirmam Berman e Ravid (2009: 96), segundo as quais, em diferentes línguas e tipos de textos, o número médio de palavras por oração cresce em função da idade/escolaridade e do registo, em particular do registo argumentativo. Corrobora-se tal afirmação no estudo de Johansson (2009), cuja pesquisa, centrada no desenvolvimento linguístico de crianças e adolescentes suecos do 4.º ano de escolaridade até ao primeiro ano da universidade, revela que a extensão média da oração contabilizada em palavras se correlaciona positivamente com o avanço nos anos escolares

nos dois registos (narrativos e argumentativos), notando-se no registo argumentativo a mudança mais expressiva. Schleifer (2003), a partir da frequência média de palavras por oração em textos de crianças e adolescentes inscritos no 4.º, 7.º e 11.º anos de escolas israelitas, encontra diferenças significativas no desenvolvimento da extensão de orações apenas no registo argumentativo, não identificando o mesmo nos textos narrativos. Ravid (2004), ao comparar textos expositivos e narrativos escritos por alunos monolíngues falantes de hebreu inscritos também no 4.º, 7.º e 11.º anos de escolaridade, também deteta que as orações dos textos argumentativos, em termos de quantidade de palavras, tendem a ser mais longas do que as dos textos narrativos, mantendo-se uma frequência crescente ao longo dos anos escolares.

Embora não se tenha identificado nenhum tipo de correlação entre a frequência média de palavras por oração e a progressão escolar, não quer isto dizer que a extensão da oração produzida pelos alunos falantes nativos de português, no registo de texto argumentativo, se mantenha inalterada desde o ano inicial até ao ano final. Como ficou ilustrado (Gráfico 6), o 7.º ano apresenta uma média de palavras por oração superior à dos outros anos, o que também foi confirmado pelo teste de Kruskal-Wallis, que apontou para diferenças substanciais entre as médias identificadas no 5.º ano quando comparado com o 7.º ano, e também entre as médias comparadas do 5.º ano e do 10.º ano, não havendo, contudo, diferenças entre as médias identificadas no 7.º e no 10.º ano. Disto se pode concluir, portanto, que é no 7.º ano que há um ponto de rutura da suposta linearidade ascendente quanto a esta medida, o que satisfaria a convicção do crescimento das unidades em função da escolaridade. Dito doutro modo, a ausência de correlação entre a frequência média de palavras por oração e a progressão escolar, nos textos estudados, parece ser amplamente afetada pela ausência de linearidade que se nota entre os grupos dos alunos estudados.

Quanto à extensão da unidade-t, os resultados aqui identificados apontam para mudanças significativas apenas no registo narrativo e, mais uma vez, opõem-se aos resultados de alguns estudos, de que se pode citar, por exemplo, Crowhurst e Piche (1979, cf. Crowhurst 1980), que, analisando a qualidade da escrita de crianças norte-americanas do quarto, sexto e nono ano, concluem que, no registo argumentativo, há uma produção mais significativa de palavras por unidades-t do que no registo narrativo. Também Nippold *et al.* (2005), num estudo que analisa a produtividade sintática de crianças, adolescentes e adultos

(respetivamente, 11, 17 e 24 anos), reportam um crescimento constante na extensão da unidade-t em textos argumentativos. Em Beers e Nagy (2009), que estudam textos narrativos e argumentativos de crianças no 7.º e no 10.º ano da escolaridade americana, identifica-se um crescimento contínuo de palavras por unidade-t, em particular mais forte nos textos argumentativos. Johansson (2009) reporta resultados semelhantes aos de Crowhurst e Piche (1979), Nippold *et al.* (2005) e Beers e Nagy (2005). Ainda quanto à extensão da unidade-t em palavras, é importante lembrar que os valores aqui identificados para a medição da extensão são matematicamente diferentes apenas entre o 5.º e o 10.º ano e entre o 7.º e o 10.º, não havendo diferenças significativas entre o 5.º e o 7.º ano.

Sobre a extensão da oração ou da unidade-t em palavras, não foram identificadas investigações focadas na escrita escolar em português de populações semelhantes à que aqui se analisa. Mais recorrentes são os trabalhos sobre a produção linguística de crianças na primeira infância, como fazem Pereira (2008), Rodrigues (2008) ou Mendes (2015). No entanto, duas investigações são notáveis. Carvalho (1989), num estudo transversal com setenta crianças entre os 7/8 e os 13/14 anos de idade, avalia a evolução da escrita nos registos narrativo e descritivo, a partir de uma perspetiva psicolinguística. Um dos resultados aponta para “uma relativa estabilidade” da extensão do período e da frase nos textos narrativos (Carvalho 1989, cf. Carvalho 2003: 35), sendo possível detetar mudanças apenas nos textos descritivos. Costa (2010), como referido no capítulo anterior, a fim de examinar o uso de conetores de valor concessivo em produções textuais dos alunos do quarto, sexto e nono anos da escolaridade portuguesa, emprega, como estratégia complementar, a medição da extensão com base na frequência média de palavras por enunciado (frase), observando que é possível notar um “crescimento entre os escritores mais jovens e os escritores adultos”, mas, semelhantemente ao que se identifica aqui quanto à extensão da oração e da unidade-t, há “uma clivagem entre os alunos mais novos, do quarto ano e do sexto ano, cujos textos não revelam aumento da extensão média de enunciados, e os alunos do nono ano” (2010: 136).

Por fim, como se comprovou com a análise estatística, é a terceira medida selecionada para avaliar a extensão da complexidade sintática – a frequência média de orações subordinadas por unidade-t – que se mostra um indicador de fortes mudanças ao longo da progressão escolar, porém, e de novo, apenas nos textos narrativos, contrariando a expectativa comum de que nos textos argumentativos o índice de ocorrência de orações por unidade-t seria

mais expressivo nos textos argumentativos, como se pode ver, por exemplo, em Johansson (2009) ou em Beers e Nagy (2009). É de notar, porém, que a extensão, em termos de uso de orações subordinadas por unidades-t, não se confirma entre o 5.º e o 7.º ano da escolaridade dos grupos investigados. A ausência de diferenças entre o 5.º e o 7.º ano de escolaridade começa a configurar-se como um padrão de ocorrências neste estudo (cf. os resultados dos testes estatísticos aplicados à complexidade lexical, tópico do capítulo anterior).

Quanto ao registo como função da progressão escolar, os estudos acima referidos para efeitos de comparação ecoam, direta ou indiretamente, a seguinte afirmação de Crowhurst e Piche (1979, cf. Beers e Nagy 2011: 188) para justificar os resultados mais expressivos centrados no registo argumentativo: “*narration places fewest demands and argument greatest demands on writers to make use of their syntactic resources*”. Numa lógica reversa, o facto de os textos narrativos serem pouco exigentes aos escritores pode, contudo, servir a uma tentativa de justificar os resultados aqui identificados, que, como se viu, não seguem na mesma direção dos resultados dos estudos citados. Deste modo, é possível especular que, dada a contínua e sistemática exposição aos textos narrativos que se vê ao longo da progressão escolar, por um lado e, por outro, uma inclusão do registo argumentativo que se pode chamar de mais cautelosa nas práticas de sala de aula, corroborada pelas proposições do Programa de Português do Ensino Básico (Portugal 2009), a instrumentalização dos recursos vocabulares (e consequentemente sintáticos) para a produção de orações e unidades-t que as crianças e os adolescentes possuem parece ser bem menos acionada nos textos argumentativos do que nos textos narrativos.

A segunda dimensão da complexidade sintática aqui avaliada – a variação –, que agrega uma leitura mais específica aos resultados da extensão, e em particular da extensão da unidade-t em orações, sustenta-se, neste estudo, nos rácios de subordinação, os quais se obtêm pela proporção de ocorrência dos quatro grandes tipos de orações subordinadas (completivas, adverbiais, relativas e gradativas) em relação ao total de orações subordinadas. Quanto a isto, os resultados da aplicação dos testes estatísticos permitem concluir que do 5.º ao 10.º ano da escolaridade há um padrão de comportamento que se manifesta mais ou menos do mesmo modo em ambos os registos, a começar pelas orações completivas, que discursivamente contribuem para a construção do texto por descreverem proposições (um facto, uma atividade, um estado, etc.) (Dixon 2006: 15). De modo idêntico nas composições

narrativas e argumentativas, a proporção de completivas não mostra mudanças significativas em relação à progressão escolar, confirmando-se assim uma das hipóteses, segundo a qual a proporção deste tipo de oração se mantém regular ao longo dos anos escolares. É interessante salientar que, nos testes de variância, a ausência de correlação pode justificar-se na inexistência de diferenças entre as médias do 5.º e do 7.º ano.

Não se confirma, porém, a hipótese levantada sobre as orações adverbiais, que, do mesmo modo que os advérbios, adicionam à oração principal uma gama vasta de significados, que incluem a localização (espacial ou temporal), a causalidade ou a finalidade, por exemplo. Especulou-se que a proporção de ocorrências deste tipo de orações se manteria estável ao longo dos anos, o que não se comprovou pelo teste estatístico, que aponta para um decréscimo significativo (correlação negativa) na proporção de orações adverbiais em relação ao total de orações subordinadas, em ambos os registos.

Uma hipótese que se confirma é referente ao crescimento dos rácios de orações relativas, em relação às quais já foi identificada uma correlação positiva com a progressão escolar, tanto nos textos de registo narrativo, como nos textos de registo argumentativo. Isto quer dizer que, conforme o aluno avança nos anos escolares, mais orações relativas são incorporadas na sua expressão linguística. Neste caso, as médias de cada grupo diferenciam-se, reforçando o seu carácter de potencial indicador de desenvolvimento linguístico.

Por fim, quanto à ocorrência de orações gradativas, a hipótese estabelecida confirma-se apenas parcialmente, visto que, apesar de haver um crescimento na proporção destas orações no registo narrativo, não há correlação alguma entre este tipo de oração e a progressão nos anos escolares.

Na tentativa de comparar resultados – salvaguardadas as devidas diferenças metodológicas –, é fundamental referir o trabalho de Hunt (1965), pois que, embora já se tenham passado cinquenta anos desde a sua publicação, é até hoje uma das principais referências na literatura sobre o desenvolvimento linguístico em idade escolar. Hunt analisa um conjunto de textos escritos por 54 crianças pertencentes a três anos distintos de escolaridade (4.º, 8.º e 12.º) e, quanto ao uso de orações subordinadas, deteta um crescimento significativo do uso de orações completivas e relativas ao longo dos anos estudados. Sobre o uso de orações adverbiais, em que se incluem as gradativas, o autor refere a inexistência de correlação com

a progressão escolar. As descobertas de Hunt (1965) confirmam-se por algumas investigações posteriores, de que se podem citar O'Donnell *et al.* (1967), Blount *et al.* (1968), Dixon (1970), Dittmer (1971), Crowhurst e Piche (1979), Perera (1984, 1986), Verhoeven *et al.* (2002), Allison *et al.* (2002), Kupersmitt (2004) e Nippold *et al.* (2005).

Entre estes, Kupersmitt (2004: 424), investigando o desenvolvimento linguístico em textos narrativos produzidos por um grupo de crianças (4, 7 e 11 anos de idade) e de adultos, distinguidos por grupos monolíngues (hebreu) ou bilingues (hebreu e espanhol), refere um crescimento significativo no emprego de orações relativas. Resultado similar se vê em Nippold *et al.* (2005). Focadas no desenvolvimento linguístico na escrita de textos argumentativos produzidos por crianças, adolescentes e adultos (11, 17 e 24 anos de idade), estas autoras também reportam um crescimento gradual no emprego de orações relativas em função da idade, mas não identificam o mesmo em relação às completivas e às adverbiais. Com base nos seus resultados, as autoras (2005: 133) sugerem que as orações relativas fazem parte de um curso desenvolvimental mais dilatado do que se vê nos outros tipos de oração.

Para Berman (2007: 353), apesar de os mecanismos de combinação de orações serem percebidos desde cedo na escrita de crianças, é apenas um aprendente em estágios mais avançados, detentor de outros conhecimentos linguísticos, como, por exemplo, um léxico mais expandido, que é capaz de os utilizar com mais propriedade. Segundo esta autora (2007: 353), tais construções são a marca registada de um estilo mais académico, mais letrado, que só tem lugar a partir do ensino secundário. Esta capacidade de empregar mais orações subordinadas – aqui delimitadas pelos domínios da unidade-t –, e mais diversificadas, revela-se, portanto, uma característica-chave do desenvolvimento linguístico em idade escolar, e em particular do desenvolvimento sintático. Tal capacidade deve ser entendida, na defesa que Slobin faz (1985, cf. Berman 2007: 354), como parte de um processo de reorganização funcional dos usos linguísticos, que se circunscreve pela atribuição de novas funções a estruturas já adquiridas e, ao mesmo tempo, pela utilização de novas formas para satisfazer funções já dominadas. Esta conceção permite explicar os modos de emprego da subordinação que se notam na construção de complexos oracionais impressos nos textos dos alunos em estágios mais avançados.

5.7. Síntese

Em consonância com o objetivo central desta tese, reportou-se, neste capítulo, um estudo de correlação estatística entre um conjunto de indicadores de complexidade sintática e a progressão nos anos escolares em dois registos (narrativo e argumentativo) produzidos por crianças e adolescentes monolíngues de português europeu. Os indicadores de complexidade sintática utilizados descreveram-se tanto em função da extensão, medida pela frequência média de palavras por oração, pela frequência média de palavras por unidade-t e pela frequência média de orações subordinadas por unidade-t, como em função da variação, medida pelo rácio de orações subordinadas (completivas, adverbiais, relativas e gradativas) pelo total de orações.

Quanto à frequência média de palavras por oração, a aplicação dos testes estatísticos revelou que este indicador de extensão sintática não mantém correlação com a progressão ao longo dos anos escolares em nenhum dos registos analisados. Embora o número médio de palavras por texto cresça do 5.º ao 10.º ano – com o reparo de que, nos textos de registo narrativo, o crescimento é linear, enquanto nos de registo argumentativo há um ligeiro decréscimo no 7.º ano em comparação com o 5.º ano –, o teste de correlação mostrou que a distribuição, quando limitada ao escopo da oração, é matematicamente equivalente nos três anos escolares em ambos os registos.

Ainda quanto à frequência média de palavras, mas desta vez limitada pelos domínios da unidade-t, verificou-se que esta medida de complexidade da extensão sintática mantém uma correlação positiva com a progressão escolar apenas no registo narrativo, não sendo identificado nenhum tipo de correlação no registo argumentativo. Quer isto dizer que, nos textos narrativos, a unidade-t, enquanto unidade de análise de textos de aprendentes para substituir a frase (Hunt 1965, 1970), torna-se mais robusta em termos de quantidade de palavras ao longo da progressão escolar.

Por fim, a aplicação de testes estatísticos revelou que a terceira medida de avaliação da complexidade sintática, que é a frequência média de orações subordinadas por unidade-t, mantém uma correlação positiva com a progressão escolar, mas unicamente nos textos de registo narrativo, não havendo correlação com os textos de registo argumentativo. Ao comparar-se este resultado com o da medida anterior, concluiu-se que, nos textos narrativos, a quantidade de orações é um indicador de mudanças da extensão sintática realizada nos textos dos alunos estudados.

Da aplicação dos testes de correlação para avaliar a variação sintática, medida pelo rácio de orações subordinadas por tipo (completiva, adverbial, relativa e gradativa) pelo total de orações subordinadas, identificou-se um cenário semelhante nos textos de registo narrativo e de registo argumentativo, diferenciando-os, em termos estatísticos, as ocorrências de orações gradativas. Em ambos os registos, não se revelou nenhuma correlação entre o rácio de orações completivas e a progressão escolar, implicando isto que não há nem aumento nem diminuição no seu uso ao longo dos anos escolares. Da comparação entre o rácio de orações adverbiais, pelo contrário, resultou uma correlação negativa, ou seja, diminui o uso deste tipo de subordinada do 5.º para o 10.º ano. É apenas o uso das orações relativas que mantém uma correlação positiva com a progressão nos anos de escolaridade, sendo a força da correlação maior no registo narrativo do que no argumentativo. Com uma expressão numérica parca, o rácio de orações gradativas, no registo argumentativo, mantém com a progressão escolar uma correlação positiva, embora fraca. Já no registo narrativo, não se vê o mesmo, não sendo identificada correlação alguma entre este tipo de oração subordinada e a progressão nos anos escolares.

6. Estudos de casos⁴⁷

Como observam Gillis e Ravid (2009: 229), as mudanças por que passa a língua de uma criança ou adolescente não se configuram como fenómenos linguísticos isolados, mas sim relacionados com complexas transformações de natureza cognitiva, social, afetiva, comportamental e, acrescento, biológicas, as quais direta ou indiretamente podem justificar as diferenças identificadas nos textos escritos escolares estudados. Justificações destes tipos, no entanto, vão muito além do escopo desta tese. Sem o mesmo rigor estatístico da descrição dos indicadores como se fez nos estudos de correlação, mas fornecendo informações de valores percentuais, apresentam-se a seguir alguns breves estudos de casos, em que se descrevem usos que os participantes deste estudo fazem de determinados recursos linguísticos, tanto nos textos narrativos, como nos textos argumentativos. Portanto, o que se segue, enquanto suplemento aos estudos estatísticos de correlação, quantitativos pela sua natureza, é uma tentativa de descrever qualitativamente casos de uso que se revelaram expressivos no processo de análise dos textos.

É importante salientar que, nestas descrições, já não se faz a mesma distinção que se aplicou, com fins metodológicos, entre os indicadores de complexidade lexical e os de complexidade sintática, dado que neste capítulo se lida com casos de uso efetivos da língua, em que os limites entre o estrato lexical e o sintático se tornam mais lábeis. Adotar tal posicionamento implica alinhar este estudo com o reconhecimento de que o léxico e a sintaxe são aspectos fortemente interdependentes quando se examinam padrões linguísticos (Butler 2003: 482). Pretende-se, assim, contribuir para uma compreensão mais abrangente dos modos como crianças e jovens dispõem da língua escrita ao longo do desenvolvimento escolar. Por não serem estudos que se esgotam em si mesmos, servem também como pontos de partida para investigações futuras.

⁴⁷ É importante reiterar que o emprego da expressão “estudos de caso” segue a aceção apresentada por Ponte (2006: 106): “É uma investigação (...) que se debruça deliberadamente sobre uma situação específica que se supõe ser única ou especial, pelo menos em certos aspectos, procurando descobrir a que há nela de mais essencial e característico e, desse modo, contribuir para a compreensão global de um certo fenómeno de interesse.”

De um modo geral, os casos apresentados a seguir, ora limitados a um registo específico, ora englobando os dois, justificam-se por ocorrências numericamente expressivas. Os casos descritos nas secções 6.1 a 6.5 originam-se nas listas (e, por vezes, sublistas) de palavras mais frequentes (ou a elas associadas) por ano e por registo (Tabela 7 e Tabela 8). Listas de palavras como estas permitem a identificação de diferenças numéricas das ocorrências que podem ser tomadas como mais relevantes (Baker *et al.* 2006: 136). As secções 6.6 a 6.8 descrevem usos de recursos linguísticos que se mostraram numericamente significativos (para mais ou para menos) durante a testagem estatística.

Vejam-se, primeiramente, as listas de palavras mais frequentes (por *tokens*) nos textos de registo narrativo e registo argumentativo, respetivamente:

Ranking	5.º ano		7.º ano		10.º ano	
	Palavra	Frequência	Palavra	Frequência	Palavra	Frequência
1	fomos	36	fomos	64	dia	55
2	ir	35	dia	64	verão	55
3	dia	34	casa	64	ir	48
4	casa	32	amigo	53	casa	43
5	praia	30	praia	50	praia	37
6	verão	26	verão	47	amigo	36
7	amiga	22	amiga	41	amiga	34
8	disse	21	ir	30	noite	31
9	férias	17	tinha*	26	tinha	31
10	ver	15	férias	20	amigos	27

Tabela 7: Lista das dez palavras lexicais mais frequentes (*tokens*) no registo narrativo ao longo da progressão escolar.

Ranking	5.º ano		7.º ano		10.º ano	
	Palavra	Frequência	Palavra	Frequência	Palavra	Frequência
1	redes	60	redes	185	redes	222
2	sociais	56	sociais	171	sociais	208
3	facebook	51	pessoas	106	pessoas	139
4	pessoas	47	facebook	70	amigos	63
5	coisas	30	importantes	53	facebook	56

6	acho	29	amigos	51	pessoa	34
7	amigos	27	dia	49	importantes	31
8	internet	27	acho	43	vida	31
9	falar	24	coisas	35	acho	30
10	fazer	24	falar	33	falar	28

Tabela 8: Lista das dez palavras lexicais mais frequentes no registo argumentativo ao longo da progressão escolar.

Segundo Berber-Sardinha e Shimazumi (2003: 18), listas de palavras podem ser indicativas do tema, ou temas, mais recorrentes num dado *corpus*. Assim, da lista apresentada na Tabela 7, cujas palavras, em valores percentuais, correspondem, no 5.º ano, a 5,99% do total de palavras; no 7.º ano, a 5,56%; e, no 10.º ano, a 3,75%, pode-se inferir claramente que a seleção lexical dos participantes reflete o tema da tarefa de escrita, que, em síntese, pede a narração de uma aventura no último verão, de que devem tomar parte o narrador e o melhor amigo. Pode-se inferir ainda desta lista de palavras que tal aventura acontece com o narrador acompanhado por um ou mais amigos, num dia das férias de verão e numa casa na praia, elementos que, do ponto de vista discursivo, tipificam a construção do enquadramento introdutório de narrativas (*introductory setting elements*) (Berman 2004: 275). É importante relembrar, como se defendeu no capítulo sobre as definições operacionais, que as tarefas de escrita são entendidas aqui como o principal contexto de situação de textos escolares, ou, como sugiro, a tarefa é o contexto de situação local (em oposição ao contexto de situação global, não-linguístico).

No registo argumentativo, como se pode ver na Tabela 8, as dez palavras mais frequentes do 5.º ano correspondem a 9,69% do total de palavras (*tokens*); no 7.º ano, corresponde a 11,29%; e, no 10.º ano, a 8,85%. E, semelhantemente ao que acontece no registo narrativo, a seleção lexical nos textos argumentativos prende-se fortemente com a tarefa de escrita, mas, diferentemente do caso anterior, a relação entre as palavras mais frequentes e a tarefa, que pede um texto de opinião sobre as redes sociais, não se realiza apenas por meio de potenciais associações semânticas, mas também por meio de repetição explícita de itens que constavam do enunciado da tarefa que era solicitada aos alunos, o que se exemplifica pela ocorrência dos nomes ‘redes’, ‘sociais’, ‘facebook’, do adjetivo ‘importantes’ e do verbo ‘achar’, que constam da tarefa.

6.1. Estatuto lexical do verbo ‘ir’

A lista de palavras mais frequentes identificadas no registo narrativo (Tabela 7), grosso modo, revela fortes semelhanças entre os três anos da escolaridade quanto ao léxico selecionado, em particular no que respeita ao modo como os alunos se vinculam à tarefa. Há, no entanto, algumas ocorrências que, porque podem diferenciar um ano de outro, merecem atenção mais pormenorizada, a começar pela recorrência do verbo ‘ir’ nos textos narrativos, que está representado, enquanto verbo lexical, nas listas dos três anos de escolaridade pelos *types* ‘fomos’ e ‘ir’. Assim, apresenta-se, na Tabela 9, abaixo, a frequência específica das dez principais formas do verbo ‘ir’ nos textos narrativos, independentemente do seu estatuto lexical, se verbo pleno ou auxiliar:⁴⁸

Ranking	5.º ano		7.º ano		10.º ano	
	Palavra	Frequência	Palavra	Frequência	Palavra	Frequência
1	fomos	56	fomos	135	fomos	54
2	ir	38	ir	31	ir	48
3	fui	20	íamos	24	íamos	23
4	ia	12	ia	17	foram	16
5	íamos	11	fui	12	iríamos	11
6	vamos	8	vou	9	fui	11
7	vai	3	irmos	7	ia	10
8	irmos	3	foram	9	iria	4
9	íam	3	vamos	8	ido	4
10	foram	3	ido	3	irmos	4

Tabela 9: Lista das dez formas mais frequentes (*types*) do verbo ‘ir’ no registo narrativo ao longo da progressão escolar.

⁴⁸ Gonçalves (2002: 63 e seguintes) conclui, com base num conjunto de testes sintático-semânticos, que o verbo ‘ir’ é semi-auxiliar, dado que apresenta “um comportamento duplo (por um lado idêntico ao dos auxiliares; por outro, idêntico ao de verbos principais, como parecer)” (2002: 80). Sigo Raposo (2013: 1221 e seguintes), que opta por incluí-lo na categoria dos verbos auxiliares.

A utilização sintático-semântica que se faz do verbo ‘ir’ nos textos narrativos expressa-se numa das seguintes funções: a) verbo pleno, seguido de complemento preposicionado (obliquo): “No outro dia fomos para outra cidade”, “Nós também fomos à praia muitas vezes”; b) verbo auxiliar temporal seguido de verbo no infinitivo: “E eu vou levar a minha melhor amiga se puder”; c) verbo auxiliar aspetual seguido de verbo no gerúndio ou no infinitivo (neste último caso, mediado pela preposição ‘a’): “à medida que íamos andando a paisagem ia melhorando cada vez mais”, “Elas no caminho iam a contar as férias fantásticas que tiveram”. A distribuição de uso destas construções nos textos narrativos ilustra-se no Gráfico 9, abaixo:

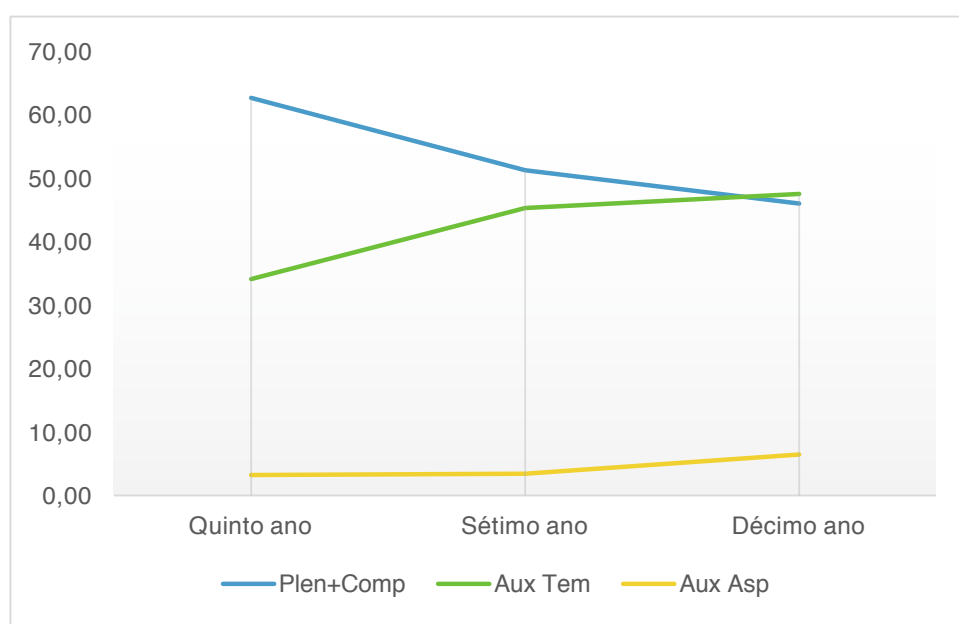


Gráfico 9: Distribuição dos valores médios de ocorrências, em percentual, de construções com o verbo ‘ir’ no registo narrativo ao longo dos anos escolares.

Como se pode ver, a distribuição de usos não é plana ao longo da progressão escolar, havendo um expressivo declínio do emprego do verbo ‘ir’ em construções Plen+Comp, o qual, no 5.º ano, representa 62,6% do total de ocorrências do verbo ‘ir’; no 7.º ano, representa 51,23% do total; e no 10.º ano, 46,04%. Ao contrário disto, os auxiliares temporais e aspetuais crescem ano após ano. No 5.º ano, 37,4% do total de ocorrências do verbo ‘ir’ ocorrem com função auxiliar (temporal ou aspetual); no 7.º ano, este valor passa a 48,77%; e no 10.º, a 53,96%.

A instanciação do verbo ‘ir’ como auxiliar nos textos narrativos contribui para a perspetivação temporal dos eventos narrados, quer através da expressão de futuridade (como

auxiliar temporal), servindo, de acordo com Cunha e Cintra (2008: 411), “para exprimir o firme propósito de executar a ação, ou a certeza de que ela será realizada em futuro próximo”, quer através da expressão da gradação do evento principal (como auxiliar aspetual), que serve, ainda segundo os gramáticos, “para indicar que a ação se realiza progressivamente ou por etapas sucessivas” (*ibidem*). Esta contribuição justifica-se porque, mesmo que este verbo seja o resultado de um processo de gramaticalização, ainda guarda a dimensão temporal do verbo pleno a que ele corresponde (Raposo 2013: 1228).

6.2. Expressão do tempo adjunto e argumental

Outro caso que merece atenção é a ocorrência significativa do item ‘dia’, que também aponta para a expressão do tempo. De acordo com Móia e Alves (2013: 557), a expressão do tempo manifesta-se em diferentes subsistemas gramaticais, sendo uma forma de expressão possível a realização adjunta, que se pode codificar com o recurso a sintagmas não-obrigatórios, podendo estes ser a) nominais (“Elas todos os dias iam para a praia de bicicleta”; b) adverbiais (“Eu ao principio assustei-me mas depois controlei-me”); ou preposicionais (“Durante o dia íamos sempre para a praia”).

Para a apresentação deste caso, restrinjo-me às ocorrências de adjuntos temporais realizados por sintagmas nominais e preposicionais com a presença explícita do item ‘dia’. Nos textos estudados, a distribuição dos sintagmas com esta função nos textos narrativos mostra-se no Gráfico 10, abaixo:

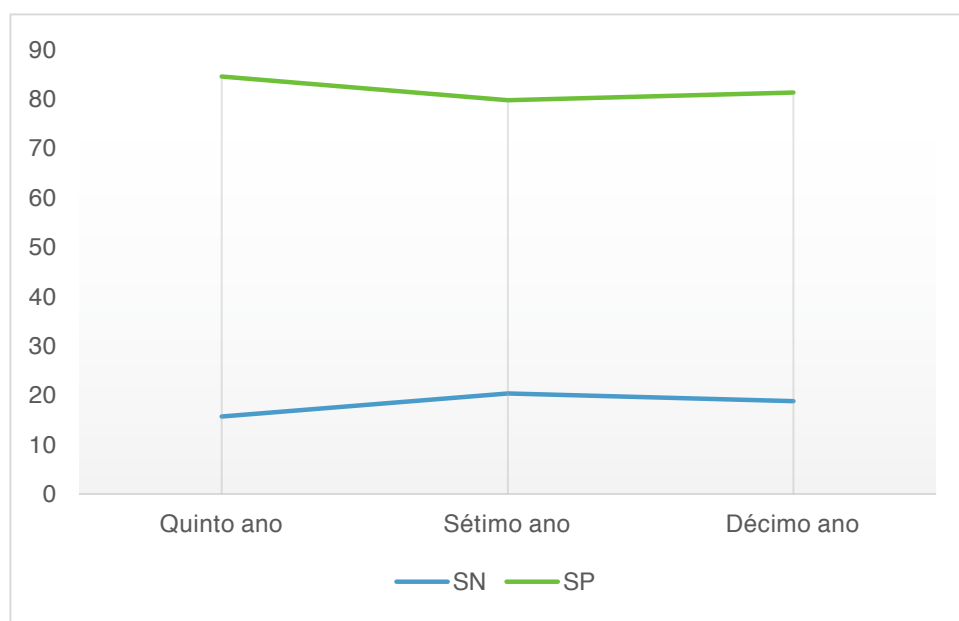


Gráfico 10: Distribuição dos valores médios de ocorrência, em percentual, de sintagmas nominais e preposicionais com a função de adjuntos temporais construídos com o item ‘dia’ no registo narrativo ao longo da progressão escolar.

Os adjuntos temporais construídos com o item ‘dia’, como se vê no Gráfico 10, são preferencialmente codificados por sintagmas preposicionais (“no dia seguinte”, “no outro dia”, “no fim do dia”, “a partir daquele dia”, etc.) e menos comumente por meio de sintagmas nominais (“um dia”, etc.), mostrando-se de uso bastante regular ao longo dos anos escolares. Em particular, nota-se que a diferença de usos é mais dilatada no 5.º ano, decrescendo ligeiramente nos anos seguintes. Nota-se, contudo, que é na seleção vocabular dos constituintes desses adjuntos que reside a variação, nomeadamente na constituição dos sintagmas preposicionais. No 5.º ano, a escolha prototípica de sintagma preposicional é “no dia seguinte”, e variações (“nos dias seguintes”, “no dia a seguir”) para expressar o desenvolvimento dos eventos narrados, o que representa 35,48% do total deste tipo de construções. No 7.º ano, esta escolha representa 17,39% e, no 10.º, apenas 10,53%.

Ainda quanto ao tempo, é relevante destacar o uso nos textos narrativos de outra construção associada à expressão do tempo, a que Mória e Alves (2013: 557) chamam de tempo argumental e que se expressa pela combinação de predicados verbais temporais com os seus argumentos. Na descrição deste caso, restrinjo-me às ocorrências dos verbos ‘chegar’, ‘passar’, ‘haver’, ‘demorar’, ‘durar’, ‘fazer’). O Gráfico 11 apresenta os valores relativos das ocorrências da expressão do tempo argumental:

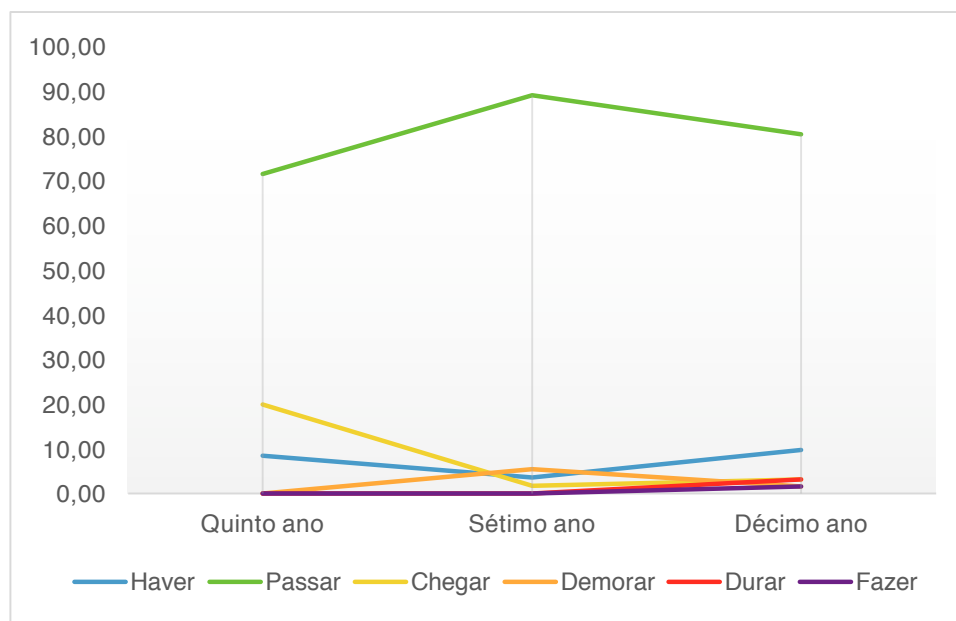


Gráfico 11: Distribuição dos valores médios de ocorrências, em percentual, de construções com predicados verbais temporais no registo narrativo ao longo dos anos escolares.

Como se pode observar, a expressão do tempo com o verbo ‘passar’ é bastante mais recorrente do que com as outras formas verbais, correspondendo, no 5.º ano, a 71,43% do total de ocorrências; no 7.º ano, este percentual eleva-se a 89,09%; e, no 10.º ano, decresce ligeiramente para 81,67%. É interessante ainda observar o gradual abandono do recurso ao verbo ‘chegar’, que, no 5.º ano, equivale a 20% do total, passando a 1,82%, no 7.º ano, e 3,28%, no 10.º ano. As outras formas verbais têm ocorrências numericamente insignificantes.

As construções que empregam o verbo ‘passar’ na expressão do tempo, em todos os anos escolares, seguem, na sua maioria, o padrão [verbo flexionado + predicado de quantidades de tempo] ou o padrão [particípio verbal + predicado de quantidades de tempo], como se exemplifica abaixo, em (5) e (6)⁴⁹, respetivamente:

5. “No verão passado foi a casa de uma amiga *passar um mês*”. (mcolfs_1_5h_mda)
6. “*Passado duas horas* ele disse-me que ia se embora...” (hsl_1_7a_atd)

Quanto ao padrão com o particípio verbal, o exemplo (6) aponta para um uso que pode ser considerado (por alguns falantes mais conservadores) desviante da norma, já que não há concordância entre o sujeito gramatical e a forma participial. Este tipo de construção é, porém, bastante recorrente em todos os anos estudados, indicando possivelmente uma mudança linguística na norma, envolvendo uma gramaticalização do verbo *passar* como conector de tipo prepositivo (Móia c.p.). No 5.º ano, do total de ocorrências, apenas numa se vê a flexão da forma participial em concordância com o sujeito (“Passados alguns minutos”), verificando-se nas restantes ocorrências a forma participial no masculino singular, independentemente do número e da pessoa do sujeito (“Passado algumas semanas”, “passado dois minutos”, “passado uma hora e meia”, etc.). O mesmo cenário se repete no 7.º e no 10.º ano.

⁴⁹ No exemplo (5), é interessante dizer que a forma verbal ‘foi’ está a ser utilizada para referir a primeira pessoa do singular, o narrador.

6.3. Influência extralinguística na variação vocabular

Ainda quanto à lista de palavras extraída dos textos narrativos, justifica-se uma breve nota sobre o item ‘noite’, presente apenas na lista de palavras mais frequentes do 10.º ano. Esta seleção exemplifica um caso de variação vocabular, não por estar relacionada com uma recategorização sintático-funcional, como se viu com o caso do item ‘ir’, por exemplo, mas por ser motivada por um fator de ordem extralinguística, que tem realização nas diferenças vocabulares que o 10.º ano sistematicamente apresenta em comparação com os outros anos, diferenças estas detetadas estatisticamente. É no 10.º ano, em que a média de idade dos alunos ronda os quinze anos, que se nota um conjunto de vocábulos não utilizados nos anos anteriores (‘discoteca’, ‘bar’, ‘copos’, ‘sexuais’, ‘embriagados’, etc.). O mesmo fator extralinguístico, isto é, a diferença etária, reflete-se igualmente na presença menos significativa de palavras referentes a elementos adultos da família (‘pai’, ‘mãe’, ‘avó’, etc.) no 10.º ano, sendo, ao contrário, itens comuns no 5.º e no 7.º ano, em que os alunos mantêm entre si diferenças mínimas de idade (10 e 12 anos, respetivamente).

6.4. Alternância no emprego de verbos epistémicos

A partir da lista das palavras mais frequentes nos textos de registo argumentativo (Tabela 8), um primeiro caso que merece ser mencionado refere-se ao emprego de alternativas ao verbo epistémico ‘achar’, que tem ocorrência significativa nos três anos estudados. Verbos de valor modal epistémico implicam “graus de certeza ou avaliação de probabilidade acerca do conteúdo proposicional da frase” (Oliveira e Mendes 2013: 623). O verbo ‘achar’ enquadra-se no primeiro caso, já que expressa, nos textos argumentativos investigados, o grau de comprometimento do sujeito da frase (ou do escritor) em relação à importância da existência das redes sociais. Outros verbos que se podem considerar epistémicos são ‘julgar’, ‘crer’, ‘pensar’, ‘calcular’, ‘imaginar’ e ‘considerar’. Estes verbos foram identificados nos textos argumentativos, a fim de comparar a sua frequência com a frequência de ‘achar’, revelando potenciais indicadores do desenvolvimento lexical associado aos valores de modalidade expressos por meios verbais. O Gráfico 12 demonstra comparativamente a utilização destes verbos ao longo da progressão nos anos escolares:

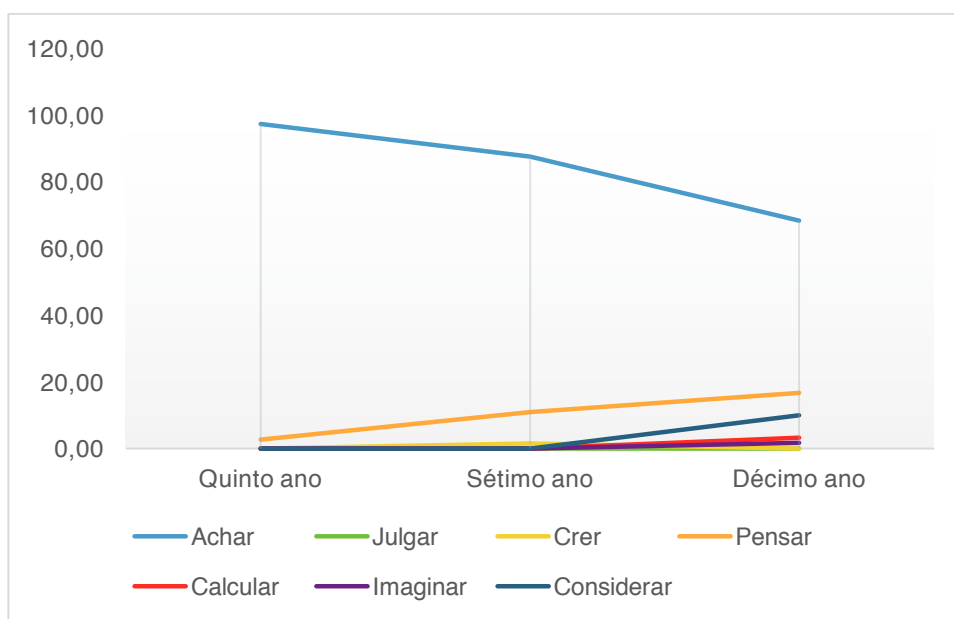


Gráfico 12: Distribuição dos valores médios de ocorrências, em percentual, de verbos modais epistêmicos no registo argumentativo ao longo dos anos escolares.

Nos textos argumentativos, como se pode inferir da lista de palavras mais frequentes e comprovar pelo Gráfico 12, a escolha prioritária para a expressão da crença, em todos os anos de escolaridade, é o emprego do verbo ‘achar’, que, no 5.º ano, representa 97,3% das escolhas, no 7.º, 87,5%, e, no 10.º ano, 68,33%. Estes valores percentuais decrescentes revelam também um acesso gradual a outras formas verbais possíveis para a expressão da modalidade epistêmica, de que se destaca o verbo ‘pensar’, que no 5.º ano equivale a 2,7% do uso total, no 7.º ano, a 10,94% e, no 10.º ano, a 16,67%. Note-se ainda a ocorrência de ‘considerar’, apenas no 10.º ano, em que representa 10% das ocorrências de verbos deste tipo.

6.5. Construções com o verbo ‘achar’

Quanto ao emprego específico do verbo epistêmico ‘achar’ no registo argumentativo, também se nota uma mudança do ano escolar inicial para o ano final, mas em relação à constituição das construções de que faz parte, as quais podem ser tipificadas por a) Achar + Que (completiva finita), como em “... acho que as redes sociais são uma coisa muito boa...”; e b) Achar X + Predicativo (ou + predicação secundária, como prefere Raposo

(2013: 1340)), como em “Eu sou contra as redes sociais, pois acho-as perigosas...”. O gráfico abaixo ilustra a distribuição dos usos destas construções ao longo da progressão escolar:

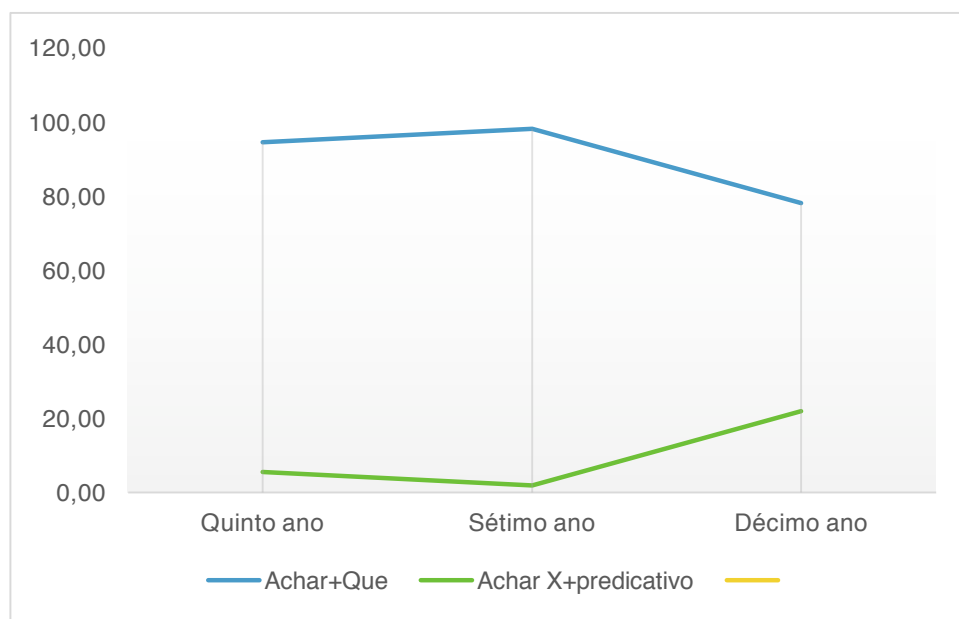


Gráfico 13: Distribuição dos valores médios de ocorrências, em percentual, de construções com o verbo epistêmico ‘achar’ no registo argumentativo ao longo dos anos escolares.

Como se vê no Gráfico 13, no decorrer da escolaridade, há, por um lado, uma tendência de movimento decrescente do uso da construção de valor epistêmico Achar + Que, sendo representada, no 5.º ano, por 94,44% do emprego do verbo epistêmico ‘achar’; no 7.º, por 98,18%; e, no 10.º ano, por 78,05%. Por outro lado, é crescente o uso da construção correlata Achar X + Predicativo, com valores percentuais de 5,56%, no 5.º ano, passando a 1,82% no 7.º ano e elevando-se sensivelmente para 21,95% no 10.º ano de escolaridade.

6.6. Incidência de pronomes pessoais e demonstrativos

Outra frequência de uso que é relevante descrever é a de pronomes com função anafórica, isto é, pronomes cuja referência é fixada pela sua ligação a outras expressões referenciais (geralmente sintagmas nominais). O estudo deste uso justifica-se não porque a ocorrência destes itens nos textos do *corpus* é numericamente expressiva, mas porque se enquadra num tópico de investigação recorrente na literatura sobre o desenvolvimento linguístico de crianças e adolescentes, a saber, a coesão textual (cf. McCutchen e Perfetti 1982;

Bamberg 1984; Fitzgerald e Spiegel 1986; Figueiredo 2000; Crossley *et al.* 2013). Como se faz nos outros estudos de caso, o objetivo é o de apresentar, sem pretensão de exaustividade, uma breve descrição da incidência da classe dos pronomes anafóricos a fim de complementar a leitura resultante dos estudos de correlação, discutidos anteriormente.

Os pronomes selecionados para este estudo de caso englobam, primeiramente, as formas dos pronomes pessoais de terceira pessoa, tanto tónicos ('ele', 'ela' e variações e 'si') como átonos ('o', 'a', 'lhe' e variações e 'se'), os pronomes possessivos ('seu', 'sua' e variações) e os pronomes demonstrativos ('isso', 'isto', 'aquilo', 'este', 'esse' e 'aquele' e variações). Nos textos de registo narrativo, tais pronomes tipicamente reportam-se a antecedentes que designam personagens que não o narrador (cf. (7)), que, por imposição da tarefa de escrita narrativa, é também uma personagem, apesar de nem sempre esta condicionante ser atendida, limitando-se o narrador a relatar a história (cf. (8)). Tal exigência é a razão pela qual os pronomes de primeira pessoa do plural, tanto na forma tónica ('nós'), como na átona ('nos'), e ainda as formas possessivas ('nosso', 'nossa' e variações) são também anafóricos, nos textos em análise, isto é, são ligados a dois ou mais antecedentes ("split antecedents", cf., e.g., Safir 2004), em que um deles refere o narrador (cf. (9)). Vejam-se os exemplos (com pronomes anafóricos em itálico e antecedentes sublinhados):

7. No primeiro dia de verão, eu fui com a minha melhor amiga para o Algarve. No dia seguinte liguei para confirmar se *ela* ia, *ela* disse que sim e eu gostei muito porque ia ter companhia. (dsra_1_5h_mda)
8. Era uma vez duas crianças a Inês e a sua amiga Madalena, *elas* foram passar as primeiras férias sozinhas na Lagoa de Albufeira. (bamm_1_5h_mda)
9. No passado mês de julho eu e a minha melhor amiga fizemos uma viagem por varias cidades do país. Como ela tem família por várias cidades, *nós* achamos que seria divertido. (is_1_10a_fdl)

Nos textos argumentativos, o emprego de pronomes pessoais ocorre tipicamente em duas situações: ou ajudam a repor, na superfície do texto, a referência às redes sociais (parte do tópico da tarefa), o que se codifica tipicamente pela forma 'elas' (cf. (10)); ou ajudam a construir sequências narrativas (cf. (11)):

10. Na minha opinião eu acho que as redes sociais são importantes porque sem *elas* não conseguíamos comunicar com pessoas que por exemplo não tem telemóvel... (ajfcb_2_7b_mda)
11. ... acho que os: Facebook, Twitter, Windows Live Space etc... não são para as crianças. (...) Um dia vi no telejornal uma pessoa que trabalhava na SIC (entrevistador) e a filha dele estava muito mal e a mãe não sabia o que é que se passava com a filha e *ela* começou a investigar quando descobriu que foi o marido dela entrevistador da SIC... . (dsra_2_5h_mda)

Ainda a respeito dos pronomes pessoais, há casos que, embora constituam exceções, são notáveis, como se vê em (12), em que o antecedente é apenas inferível, ou em (13), em que ocorre depois do item anafórico, num movimento de recuperação prospectiva (ou catafórica):

12. Eu e a minha amiga Sofia vamos sempre à praia no verão. A praia é em Algarve. Em casa da minha avó eu e a Sofia estávamos a comer quando vimos numa notícia que um tubarão-martelo estava à solta no Algarve, pois eu e a Sofia ficamos com medo. Então cada vez que chegava a hora de ir para a praia nós escondíamo-nos sempre, porque tínhamos muito medo que o tubarão-martelo estivesse naquela praia. Mas acabaram por nos encontrar, nós dissemos que tínhamos muito medo mas *eles* levaram-nos. (basb_1_5c_mda)
13. Eram seis da manhã do dia vinte e sete de julho de dois mil e onze. Juntamo-*nos* no aeroporto de Lisboa com o fim de embarcarmos no avião com destino a Copenhaga. Ansiávamos por este momento fazia já dois anos e era ali que já se estava a concretizar. O *nosso* objectivo? Passar doze dias completamente inesquecíveis em Kristianstad, na Suécia, rodeadas de milhares de escuteiros vindos de mais de cento e cinquenta países para viver o espírito do vigésimo segundo Jamboree Mundial. Eu, a Catarina e a Beatriz juntamente com os escuteiros do nosso agrupamento ao chegar a Copenhaga sentimo-nos acolhidos com um afecto imenso por parte dos outros escuteiros. (cm1_1_10h_atd)

Vale a pena mencionar a utilização do pronome de primeira pessoa do plural nos textos de registo argumentativo, que tem ocorrência bastante expressiva. Neste registo, não exercem a mesma função anafórica que exercem no registo narrativo, motivo por que não

são considerados na frequência deste estudo de caso. Estes pronomes parecem, antes, configurar um cenário de interação entre o autor e os destinatários, em que os pronomes têm referência exofórica. Veja-se (14):

14. Hoje em dia eu acho que é importante, porque sem as redes sociais *nós* não podíamos comunicar tão rapidamente ou facilmente mas sempre com autorização dos pais, até porque nesta altura do campeonato até os *nossos* pais têm Facebook, Twitter, entre outros. (bg_2_7a_hnn)

Além dos pronomes pessoais e possessivos, também são considerados os pronomes demonstrativos ‘isto’, ‘isso’ e ‘aquilo’ e as formas variáveis ‘este’, ‘esse’ e ‘aquele’, estes últimos quer na função de pronomes, em sentido estrito, quer na de determinantes. Os demonstrativos contribuem de modo variado para a continuidade do texto: seja apontando para entidades comuns, como pessoas ou objetos (cf. (15), (16) e (17)), que tipicamente são codificadas por sintagmas nominais; seja para entidades de natureza situacional (expressas por meios proposicionais), entidades ontologicamente mais complexas, pertencendo, portanto, a uma ordem superior (e.g., Gundel *et al.* 2002: 356). Como sintetiza Asher (1993: 1), estas entidades são objetos abstratos, proposições, factos, eventos, atividades, situações, etc.⁵⁰

15. Um dia eles levaram-nos ao seu planeta os ets eram gigantes os seus cabelos eram longos e eles eram magrinhos. (...) mas eu não queria ir com eles para *aquele* planeta estranho (ag_1_10a_fdl)
16. ... nunca tínhamos visto um ser vivo daquele tamanho morto, no meio do nada. Era tão bonito mas, numa altura qualquer, num sítio qualquer e de uma forma qualquer, essa beleza tinha acabado. E ninguém sabia como. Fomos embora imediatamente a seguir, mas *isso* não importou: nosso dia estava irremediavelmente arruinado. (ba_1_10a_fdl)
17. E eu também acho que se não existissem redes sociais nós, as pessoas, não convivíamos tanto. *Esta* é a minha opinião. (bits_2_5c_mda)

⁵⁰ Nestes casos, além dos demonstrativos, atuam também como anafóricos a pró-frase átona *o*, que, neste caso, é sempre invariável, e o pronome relativo de frase *o que*.

Para a expressão da incidência dos pronomes anafóricos selecionados, diferentemente do que acontece nos outros estudos de caso apresentados nesta dissertação, em que se aplica a frequência percentual, aqui lança-se mão da frequência normalizada, em que se divide a frequência absoluta dos itens em causa pelo total de palavras (*tokens*), multiplicando-se, de seguida, por mil (Biber 1995: 405).⁵¹ O Gráfico 14 e o Gráfico 15 exibem o movimento das frequências normalizadas dos pronomes anafóricos nos textos narrativos e argumentativos respetivamente ao longo dos anos escolares:

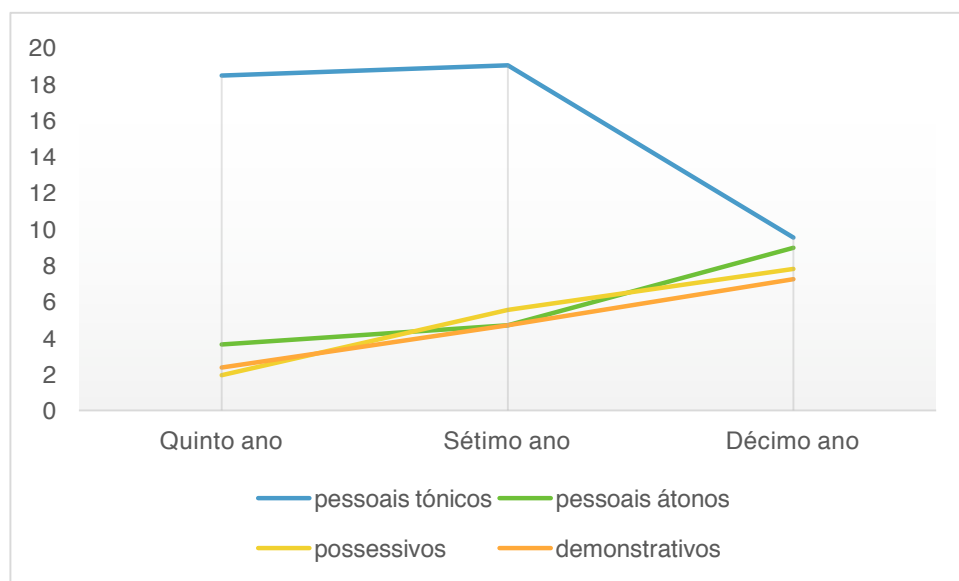


Gráfico 14: Distribuição da frequência normalizada de pronomes anafóricos no registo narrativo ao longo dos anos escolares.

⁵¹ Esta frequência assemelha-se ao que se conhece como densidade anafórica (Tutin 2002), que se mede pelo rácio entre os itens anafóricos e o total de itens – anafóricos ou não – identificados num texto.

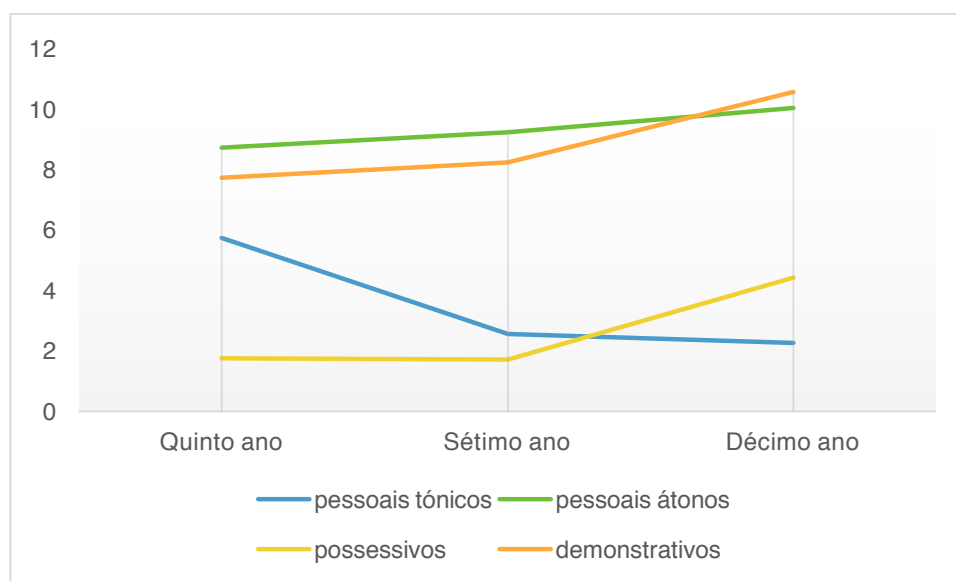


Gráfico 15: Distribuição da frequência normalizada de pronomes anafóricos no registo argumentativo ao longo dos anos escolares.

Como se pode ver, a apresentação da frequência dos pronomes pessoais tónicos faz-se separadamente da dos pessoais átonos, já que estes tipos exibem, em ambos os registos, uma dinâmica divergente: o emprego dos pronomes pessoais tónicos decresce expressivamente do 5.º para o 10.º ano, enquanto os pronomes pessoais átonos são de uso sempre ascendente. Em particular, no registo narrativo, identifica-se, no 5.º ano, uma frequência média de 18 pronomes pessoais tónicos por cada mil palavras; no 7.º, 19; e no 10.º, 10. Quanto aos pronomes pessoais átonos, no 5.º, tem-se, por cada mil palavras, uma frequência de 4 pronomes deste tipo; no 7.º ano, 5; e no 10.º, 9. No registo argumentativo, a utilização dos pronomes pessoais tónicos é, em comparação com a sua contraparte narrativa, bem menos marcante, havendo, no 5.º ano, a frequência de 6 pronomes por cada mil palavras; no 7.º, de 3; e no 10.º, de 2. Os pronomes pessoais átonos no registo argumentativo exibem, desde o ano inicial, frequências mais relevantes do que no registo narrativo, razão por que os valores são bastante uniformes em todos os anos. No 5.º e no 7.º ano, a frequência é de 9 por cada mil palavras; e no 10.º, 10.

Os pronomes pessoais possessivos, de um modo geral, comportam-se semelhantemente nos dois registos, ou seja, o seu emprego cresce do 5.º ao 10.º ano. No registo narrativo, verifica-se uma frequência de 2 por cada mil palavras; no 7.º, de 6; e no 10.º, de 8. No registo argumentativo, o crescimento de uso diferencia-se porque acontece somente a partir do 7.º ano, sendo a frequência igual à do 5.º, 2 por cada mil palavras. No 10.º ano, verifica-

se uma frequência de 4. Quanto aos pronomes demonstrativos, o seu uso é sempre mais robusto nos textos de registo argumentativo do que nos de registo narrativo. No 5.º ano, no registo narrativo, a frequência é de 2 demonstrativos por conjunto de mil palavras; no 7.º, de 5; no 10.º, de 7. No registo argumentativo, a frequência destes pronomes no 5.º e no 7.º ano é igual, 8. No 10.º ano, esta frequência passa a 11 demonstrativos por cada mil palavras.

Grosso modo, pode-se afirmar que o cenário descrito acima relativo à utilização de pronomes anafóricos identificada nos textos analisados ilustra o que dizem Biber *et al.* (2007: 238), segundo os quais diferenças de incidência de pronomes anafóricos refletem diferenças da natureza dos registos. Não se pode negar, no entanto, que a utilização crescente de pronomes demonstrativos e pessoais átonos indica um percurso de desenvolvimento que inclui o domínio destas formas de anáfora, o que pode estar associado ao domínio de recursos anafóricos de um modo geral. Também por não haver mudanças significativas no emprego das formas verbais em nenhum dos registos estudados (como comprovado pelo estudo de correlação), especula-se que o emprego sempre descendente de pronomes pessoais tónicos nos textos narrativos e argumentativos espelha a compreensão de que, no português, enquanto língua de sujeito nulo, tais recursos são prescindíveis em muitos contextos de uso.

6.7. Orações relativas

As orações relativas na produção linguística de crianças e jovens em idade escolar na realidade portuguesa têm merecido a atenção de alguns estudos (Vasconcelos 1991, 1996; Valente 2008; Antunes e Brito 2008; Costa *et al.* 2009; Choupina 2010). Choupina (2010: 52), por exemplo, conclui que a criança tem “desde muito cedo uma gramática que lhe permite gerar frases com orações relativas”. A existência do movimento crescente de ocorrências do 5.º ao 10.º ano, como comprovado pelos testes de correlação aqui aplicados, não contraria esta convicção, mas suplementa-a, já que este movimento quantitativo aponta para um processo contínuo de desenvolvimento, implicando, por hipótese, que a um dado conjunto de estruturas já adquiridas se vão incorporando novas formas de utilização ao longo da escolaridade. Perante isto, o que se impõe é saber como, nos textos do *corpus*, este tipo de recurso gramatical se modifica ano a ano, não somente pela sua tipologia fundamental (restritivas, apositivas de sintagma nominal e apositivas de frase), mas também pela seleção dos constituintes relativos, pelo papel sintático destes na subordinada e pela sua adequação

ao registo escolar escrito e formal. Para tanto, serão consideradas apenas as orações iniciadas por morfemas relativos, e em particular por *que*, excluídas, portanto, orações não finitas (por exemplo, “Um dia estava na praia, à noite e de repente do nada aparece um negro [vindo do nada]” (gs_1_10a_fdl), que inclui uma oração participial adjetiva, funcionalmente afim de uma relativa restritiva).

Em confirmação dos valores dos rácios identificados no estudo estatístico, do total de orações subordinadas identificadas nos textos narrativos do 5.º ano, 9,73% são relativas; já no 7.º ano, as relativas equivalem a 15,77% do total; e, no 10.º ano a 26,12%. Nos textos argumentativos, o emprego de orações relativas é percentualmente mais expressivo do que nos textos narrativos; porém, as diferenças entre os anos escolares, nestes textos, são menos dilatadas do que naqueles. No 5.º ano, as relativas correspondem a 26,15% do total de subordinadas; no 7.º, a 28,50%; e, no 10.º ano, a 33%. O Gráfico 16 ilustra o movimento de crescimento no uso das orações relativas ao longo dos anos escolares em ambos os registos:

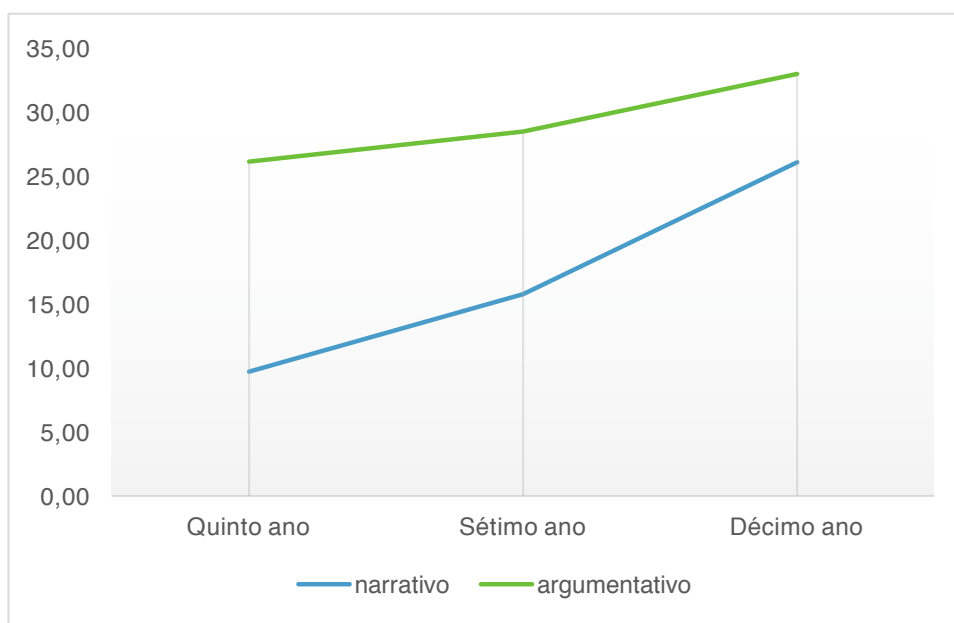


Gráfico 16: Distribuição dos valores médios de ocorrências, em percentual, de orações relativas nos registos narrativo e argumentativo ao longo dos anos escolares.

As orações relativas, como afirmam Peres e Mória (1995: 273), “permitem introduzir no discurso uma qualificação (...) acerca de uma ou mais entidades envolvidas na referência desse discurso.” A oração relativa pode ter uma função identificativa, delimitando o domínio de referência de um dado constituinte, tipicamente um sintagma nominal. Classifica-se esta função como restritiva. As orações relativas podem também ter uma função explicativa,

adicionando acerca de uma entidade – previamente identificada – uma informação que é apenas suplementar e que não é necessária para a identificação da entidade referida. Este é o tipo de oração relativa apositiva (também conhecida como explicativa). A oração apositiva pode descrever uma informação suplementar não somente sobre uma entidade identificada por um sintagma nominal, mas também sobre um conteúdo expresso por uma estrutura frásica, tratando-se, então, de uma relativa de frase.

Nos textos narrativos do *corpus*, a função identificativa das orações relativas mostra-se predominante – e sempre crescente. As relativas de frase, embora percentualmente inexpressivas se comparadas com as outras relativas, também se mostram crescentes. No 5.º ano, do total de relativas, 80% são restritivas e 20% são apositivas de SN, não havendo ocorrências de relativas de frase; no 7.º ano, as restritivas ocupam 82,28% do total, enquanto o tipo apositivo de SN se restringe a 16,46% e o de frase a 1,27%; e, no 10.º ano, 81,91% das orações funcionam como restritivas, 15,08% como apositivas de SN e 3,02% como relativas de frase.

Nos textos argumentativos, também predominam as orações relativas com função restritiva. No entanto, diferentemente do que acontece com os textos narrativos, nos argumentativos é o tipo apositivo (quer de SN, quer, mais uma vez, de frase) que cresce de ano para ano, em oposição ao tipo restritivo. No 5.º ano do registo argumentativo, do total de relativas, 98,97% são restritivas, correspondendo as apositivas de SN a apenas 1,03% e não havendo ocorrências de relativas de frase; no 7.º ano, as restritivas ocupam 95,27% do total, enquanto as apositivas de SN ocupam 3,55% e as de frase 1,18%; e, no 10.º ano, do total de relativas, 88,74 são restritivas, 7,36% são apositivas de SN e 3,90% são de frase. Os gráficos a seguir (Gráfico 17 e Gráfico 18) reproduzem a dinâmica, no *corpus* considerado, da distribuição dos três tipos de relativas descritos acima ao longo da progressão escolar, nos registos narrativo e argumentativo:

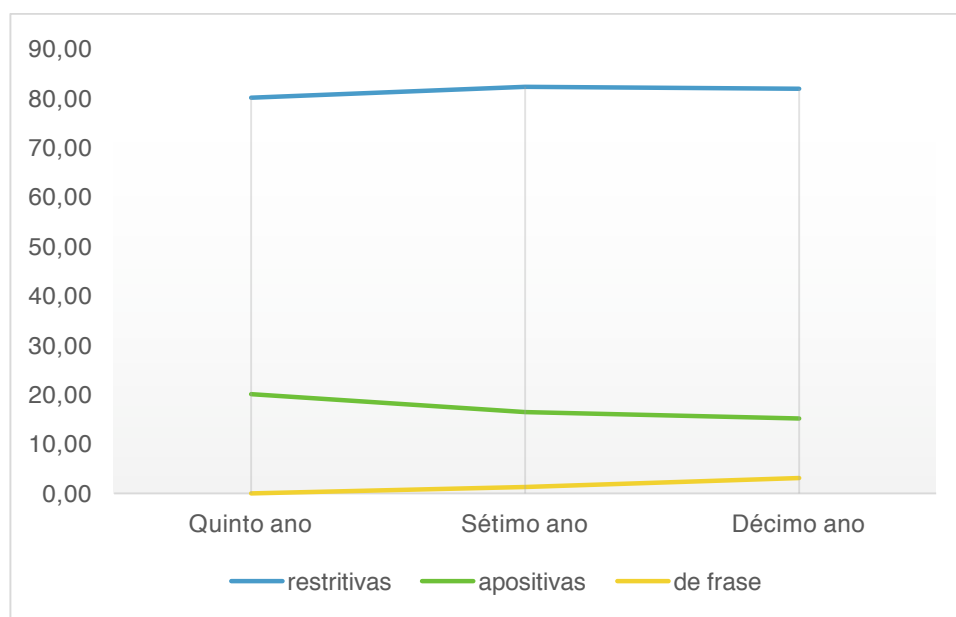


Gráfico 17: Distribuição dos valores médios de ocorrências, em percentual, de relativas restritivas, apositivas e de frase no registo narrativo ao longo dos anos escolares.

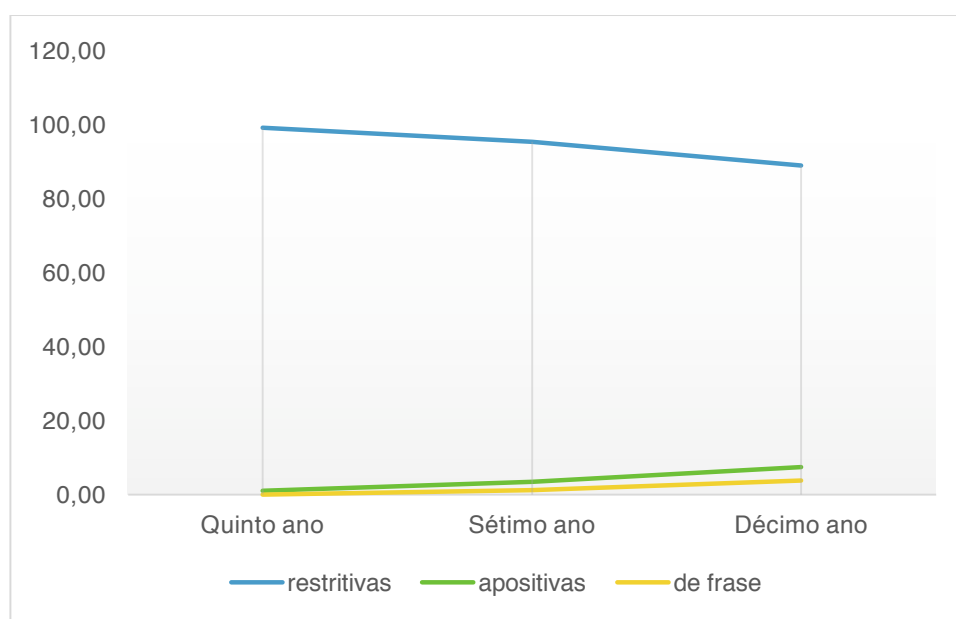


Gráfico 18: Distribuição dos valores médios de ocorrências, em percentual, de relativas restritivas, apositivas e de frase no registo argumentativo ao longo dos anos escolares.

Os dois breves estudos de caso que seguem tratam da diversidade de uso dos pronomes relativos empregues nos textos narrativos e argumentativos, com foco na função sintática que exerce na oração relativa e também na sua adequação ao registo escolar escrito e formal.

6.7.1. Diversidade de uso e função sintática de pronomes relativos

Tipicamente, todos os tipos de orações relativas são iniciados por constituintes relativos – um pronome relativo isolado ou acompanhado por outras palavras – que, além de atuarem no papel de elemento subordinante na junção entre a oração relativa subordinada e a sua subordinante, permitem retomar um dado item que está localizado na oração matriz, que é, portanto, o seu antecedente. As possibilidades de pronomes relativos, em português, dispõem-se entre as formas invariáveis ‘que’, ‘o que’, ‘quem’, ‘onde’, ‘como’, ‘quando’, ‘quanto’ e as formas variáveis ‘o qual’, ‘a qual’, ‘os quais’, ‘as quais’, ‘cujo’, ‘cuja’, ‘cujos’, ‘cujas’, ‘quanto’, ‘quanta’, ‘quantos’, ‘quantas’. Assim, anotaram-se todas as formas de pronome relativo empregues pelos alunos do 5.º, do 7.º e do 10.º ano estudados.

Em todos os anos escolares de ambos os registos, identificou-se uma utilização bastante expressiva do pronome ‘que’ (independente de estar inserido ou não em sintagma preposicional ou de ser utilizado adequadamente), embora haja um decréscimo nessa utilização do 5.º ao 10.º ano nos dois registos, sendo, no registo narrativo um decréscimo não linear, ao contrário do que se vê no argumentativo.

Mau grado uma disparidade na utilização deste recurso entre os registos, configura-se esta como a escolha maioritária para introduzir a oração subordinada relativa. Em valores percentuais, do total de pronomes seleccionados no 5.º ano do registo narrativo, 70,83% das escolhas recaí sobre o morfema ‘que’; no 7.º, 82,43%; e, no 10.º, 63,45%. No registo argumentativo, o percentual de ocorrência deste morfema, no 5.º ano, é de 76,34%; no 7.º, de 75,58%; e, no 10.º, de 64,19%. O Gráfico 19 e o Gráfico 20 ilustram o movimento de seleção de pronomes relativos ao longo da progressão escolar nos registos narrativo e argumentativo, respetivamente:

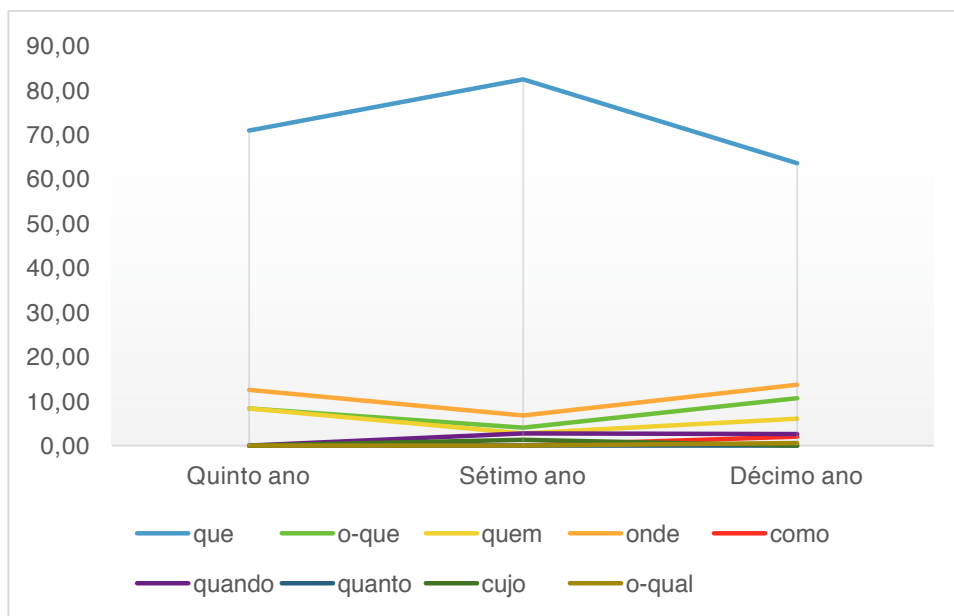


Gráfico 19: Distribuição dos valores médios de ocorrências, em percentual, de pronomes relativos no registo narrativo ao longo dos anos escolares.

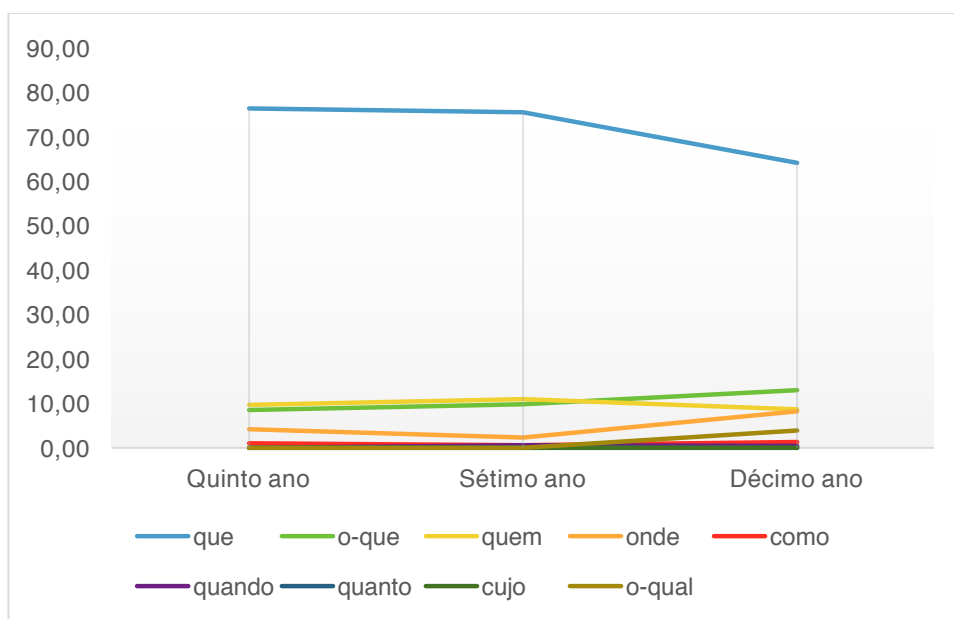


Gráfico 20: Distribuição dos valores médios de ocorrências, em percentual, de pronomes relativos no registo argumentativo ao longo dos anos escolares.

A ocorrência numericamente expressiva do morfema ‘que’ justifica uma discussão mais detalhada sobre este tipo de pronome, a começar pelas funções sintáticas que assume nas orações relativas de que faz parte. A seleção do ‘que’ como pronome relativo primário reflete a sua polivalência em termos sintáticos. Por ser semanticamente subespecificado (isto é, por poder apresentar os traços $[\pm\text{humano}]$, $[\pm\text{lugar}]$, $[\pm\text{tempo}]$, etc.), o relativo

‘que’⁵² pode exercer várias funções sintáticas dentro da oração relativa: sujeito, objeto direto ou – em conjunção com elementos preposicionais ou afins – indireto, complemento oblíquo, adjunto locativo (tempo e espaço) ou ainda modificador de nome. Nos textos narrativos analisados, é a função de sujeito a mais instanciada, ainda que sempre decrescente conforme se avança na escolaridade. No 5.º ano, equivale a 64,71% das ocorrências; no 7.º ano, a 63,93%; enquanto, no 10.º, a 54,40%. A segunda função sintática mais recorrente para o morfema relativo ‘que’ no registo narrativo é a de objeto direto⁵³), que, ao contrário do uso anterior, é sempre crescente, sendo, no 5.º ano, 11,76% do total, passando a 26,23% no 7.º, e, no 10.º ano, a 28,00%.

Do mesmo modo, no registo argumentativo, o morfema ‘que’ com função de sujeito da relativa é a escolha preferencial. Porém, o percurso da sua utilização é, ao contrário do que se vê no registo narrativo, sempre crescente. No 5.º ano, o *que* com função de sujeito corresponde a 47,89% das ocorrências, subindo para 65,38% no 7.º ano e para 67,35% no 10.º ano. O pronome relativo ‘que’ com função de objeto direto, nos textos argumentativos, também é a segunda opção. Porém a sua utilização é sempre decrescente. No 5.º ano, equivale a 36,62% das ocorrências; no 7.º, a 28,46% e, no 10.º, a 22,45%. O Gráfico 21 e o Gráfico 22, abaixo, representam a distribuição das funções sintáticas exercidas pelo relativo ‘que’ em ambos os registos, narrativo e argumentativo:

⁵² Ou antes, o constituinte dentro da frase que se segue ao pronome relativo e que é ligado por ele. Por simplicidade de exposição, uso a formulação dada no corpo do texto.

⁵³ Os casos de complemento (objeto) indireto e de complemento oblíquo, por serem numericamente inexpressivos, foram agrupados sob o rótulo “outras funções”.

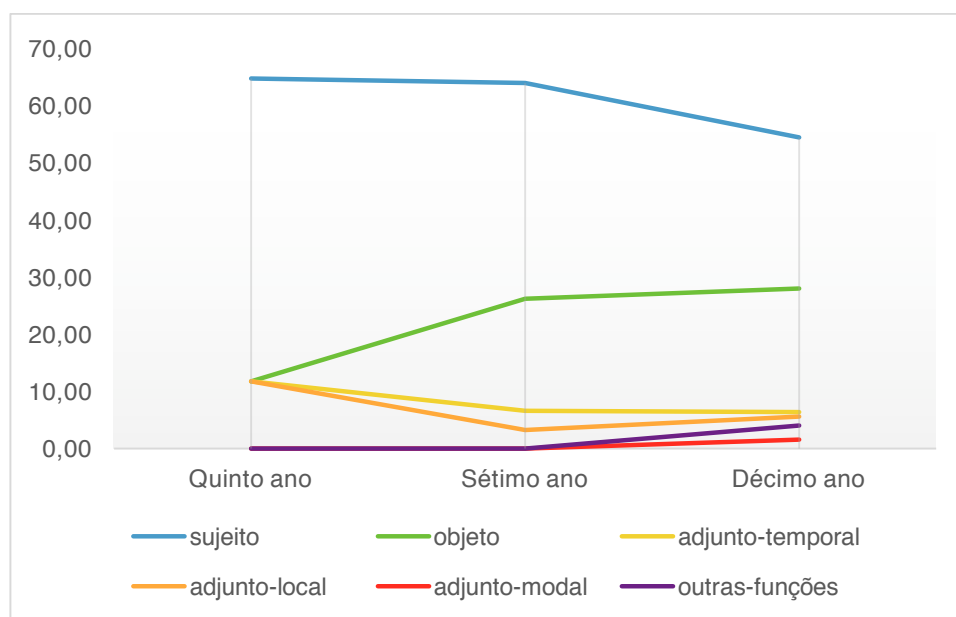


Gráfico 21: Distribuição dos valores médios de ocorrências, em percentual, das funções sintáticas exercidas pelo morfema 'que' dentro de orações relativas no registo narrativo ao longo dos anos escolares.

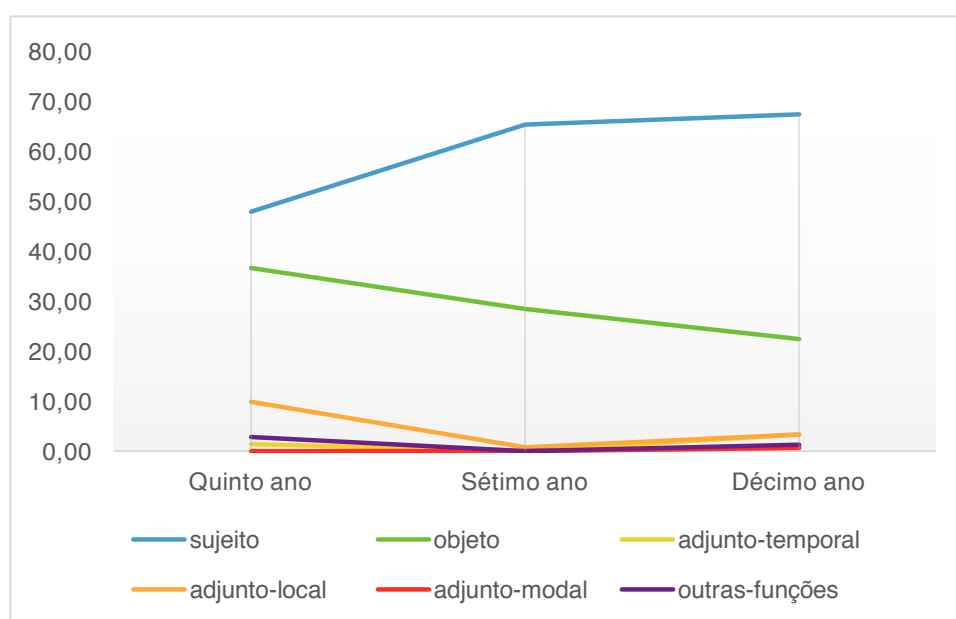


Gráfico 22: Distribuição dos valores médios de ocorrências, em percentual, das funções sintáticas exercidas pelo morfema 'que' dentro de orações relativas no registo argumentativo ao longo dos anos escolares.

O emprego ascendente ou descendente do morfema 'que', seja em posição de sujeito, seja em posição de objeto direto, parece estar relacionado com a função discursiva que a

oração relativa desempenha no texto e, consequentemente, com o registo em que se enquadra. No estudo de Jisa e Kern (1998), que avaliam textos narrativos (orais) produzidos por crianças de cinco, sete e dez anos de idade e também de adultos, reporta-se que há, da parte dos participantes mais novos, uma estratégia sintática (e lexical, acrescento) que restringe o uso de verbos transitivos. *Mutatis mutandi*, o mesmo se pode especular sobre os textos narrativos do *corpus* aqui estudado. Comprova-o, no 5.º ano, o facto de o verbo da oração subordinante relativa ser, na maioria das ocorrências, um verbo copulativo, nomeadamente ‘ser’, ‘estar’ e ‘ficar’ (cf. (18)), sendo outras formas verbais, como ‘marcar’ ou ‘assustar’, de ocorrência pontual (cf. (19)). Este cenário repete-se, com pouca variação, nos textos narrativos do 7.º ano, em que o verbo copulativo atua na atribuição de propriedades também de valor locativo (cf. Raposo 2013: 1328) (cf. (20)). A par destas ocorrências, é apenas no 7.º ano que se podem ver ocorrências do verbo leve ‘ter’ com o carácter identificativo (cf. (21)). Vejam-se os exemplos, com a relativa iniciada pelo morfema ‘que’ em função de sujeito entre parênteses retos:

18. ... tinha uma professora chamada Paula [que era um pouco má]... (gaco_1_5h_mda)
19. ... e a melhor coisa [que me marcou mais] foi a ida à Madeira. (jdld_1_5c_mda)
20. ... apaixonei-me por uma das raparigas [que lá estavam na casa da minha prima] (jx_1_7a_hnn)
21. ... tive um grande sonho com a minha melhor amiga, a Ângela Mendonça [que tem doze anos] ... (cga_1_7a_hnn)

É, de facto, no 10.º ano que se amplia sensivelmente a gama de verbos da oração relativa, havendo, portanto, mais construções com verbos transitivos, que permitem ao relativo *que*, para além da função de sujeito (cf. (22)), também funcionar como objeto (cf. (23), direto, neste exemplo). De acordo com Dasinger e Toupin (1994, cf. Jisa e Kern 1998: 630), neste uso narrativo, a oração relativa pode tornar-se uma continuadora da ação primeiramente expressa na oração principal, contribuindo para o desenvolvimento mais eficiente do enredo, como se pode ver nos exemplos (22) e (23), com os pronomes relativos em causa destacados:

22. o facto mais marcante [que se sucedeu com um grande amigo meu que fiz lá] foi o falecimento desse meu amigo. (ag_1_10a_fdl)
23. ... tivemos de ajudar nas tarefas mais básicas do campo, como limpar e colocar cloro na piscina, podar as árvores e limpar lixo [que possivelmente possamos ter espalhado pelo terreno]. (ip_1_10a_fdl)

Nos textos de registo argumentativo, cuja tarefa de escrita pedia que os alunos manifestassem a sua opinião sobre as redes sociais virtuais, há outro caso de variação vocabular que pode estar associado ao uso crescente do relativo ‘que’ com a função de sujeito, como foi demonstrado acima pelo Gráfico 22. Esta variação diz respeito ao emprego de expressões utilizadas para referir entidades (tipicamente de valor humano) que representam diferentes utilizadores das redes sociais. Estas entidades, no 5.º ano, são referidas maioritariamente com a expressão nominal ‘pessoas’ (cf. (24)), diversificando-se no 7.º (cf. (25)) e no 10.º ano (cf. (26)), em que se encontram mais reiteradamente outras expressões, como, por exemplo, ‘amigos’, ‘família’, ‘familiares’, ‘adolescentes’, ‘jovens’, ‘piratas da net’, ‘viadores’, etc., para além da expressão ‘pessoas’, que também passa a ocorrer como constituinte de sintagmas nominais mais complexos, continuando bastante presente nos textos dos alunos mais velhos. Estas expressões, em paralelo com a seleção da expressão de sentido genérico ‘redes sociais’ (ou formas de sentidos mais específicos, como ‘Google’, ‘Twitter’), ocupam maioritariamente, nas composições argumentativas, a função de antecedentes do pronome relativo ‘que’. Abaixo seguem-se exemplos dessa utilização:

24. Há pessoas [que têm família fora do nosso país]... (rtsp_2_5c_mda)
25. ... podemos travar amizade com os piratas da net [que nos podem reconhecer] ... (pp_2_7a_atd)
26. conheço casais felizes [que se conheceram nas redes sociais]. (fana_2_10h_atd)

6.7.2.(In)Adequação no emprego do pronome relativo “que”

A seleção massiva do pronome relativo ‘que’ realizada por crianças em idade escolar é também descrita em Antunes e Brito (2008). Num estudo de desenho experimental para avaliar a ocorrência de relativas restritivas com antecedente expresso produzidas por alunos

do 9.º ano da educação básica (não-longitudinal, portanto), os autores (2008: 240) afirmam que, dada a percentagem de acerto no uso deste pronome *que* (cerca de 95%, segundo eles) em relação à norma culta, as relativas iniciadas por este morfema são estruturas já consolidadas ao final do ensino básico. De facto, pelos resultados aqui encontrados – e com base num valor percentual aproximado ao referido por Antunes e Brito –, a consolidação no uso do relativo ‘*que*’ parece acontecer mesmo apenas nos anos finais da escolaridade básica, porém não se podendo, porém, generalizar tal conclusão quando se tem como variável o registo. O Gráfico 23 ilustra o movimento, em valores percentuais, da utilização adequada do pronome ‘*que*’, tomados globalmente todos os tipos de desvio, em ambos os registos:

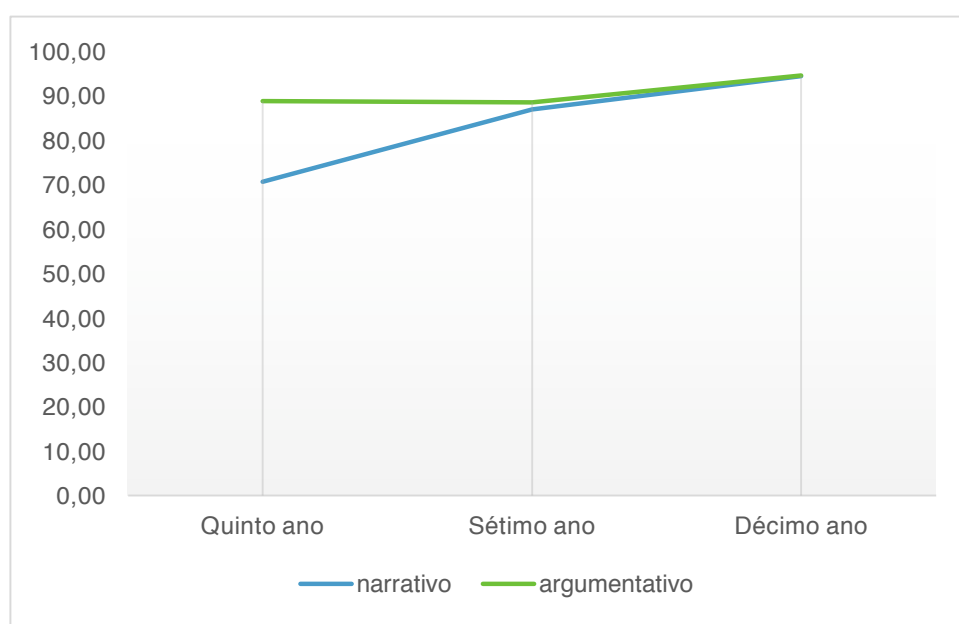


Gráfico 23: Distribuição dos valores médios de ocorrências adequadas, em percentual, do pronome relativo ‘*que*’ nos registos narrativo e argumentativo ao longo dos anos escolares.

O Gráfico acima ilustra uma disparidade no uso adequado do morfema ‘*que*’ nos textos narrativos do 5.º ano, em que 70,59% das ocorrências deste pronome relativo são conformes à norma padrão, em comparação com 88,73% de ocorrências adequadas nos textos argumentativos do 5.º ano. É apenas no 7.º ano que se dá um emparelhamento. No 7.º ano e nos textos de registo narrativo, o percentual de acerto é de 86,89%, valor muito próximo do que se encontra nos textos argumentativos do 7.º ano, com 88,46% de ocorrências adequadas. O emparelhamento mantém-se até ao 10.º ano, sendo o percentual, no registo narrativo, de 94,40% e, no argumentativo, de 94,53%, valores muito próximos dos referidos por Antunes e Brito (2008) para assegurar a consolidação no uso de relativas.

Importa referir que os desvios de adequação no uso do pronome relativo ‘que’ considerados neste estudo, a partir de Peres e Mória (1995) em particular, mas também de Brito e Duarte (2003), Antunes e Brito (2008) e Veloso (2013), refletem o que se conhece como estratégias de relativização não-canónicas. São duas as construções não-canónicas mais comuns: a eliminação de qualquer preposição obrigatória no constituinte relativo (cf. (27)), ou estratégia cortadora, e a inserção de uma pró-forma dentro da parte proposicional da oração relativa, na posição ligada pelo pronome relativo, ou estratégia de retoma (ou ressumtiva, ou ainda de duplo preenchimento) (cf. (28)). Também se consideram como desvios tanto a escolha indevida do morfema ‘que’, em detrimento de outro, mais adequado portanto (cf. (29), onde, além da omissão da vírgula, foi usado na relativa de frase, o pronome relativo ‘que’ e não a forma ‘o que’, adequada à interpretação mais natural do sentido da sequência), como também a discordância entre o pronome relativo com a função de sujeito, que comporta os valores de número e/ou género do antecedente, e o verbo da oração relativa a que se liga (cf. (30)). É de notar que esta discordância tem aceitação como variante – na gramática do português padrão – por parte de alguns gramáticos (como notam, aliás, Peres e Mória). Vejam-se os seguintes exemplos, em que o pronome relativo ‘que’ envolvido em algum tipo de desvio está sublinhado:

27. No último verão eu passei com o meu melhor amigo nós fomos para vários sítios mas houve um sítio que eu gostei mais por uma coisa que aconteceu. (bfb_c_1_7b_mda)
28. O meu melhor amigo, que não o conheci no campo, disse para eu experimentar ir para o campo... (ag_1_10a_fdl)
29. ... depois de andar nesse escorrega que foi brutal, foi me embora com o meu pai e o meu melhor amigo... (af_1_7a_hnn)
30. Uma das razões que me faz ser apologista destes meios de comunicação é o facto de me permitir entrar em contacto com os meus amigos... (ca1_2_10h_atd)

Em (27), a preposição ‘de’ está ausente, a despeito de o verbo ‘gostar’ exigir que o seu complemento seja realizado por um sintagma preposicional (comparar com *um sítio de que eu gostei mais*). Em (28), a inserção do clítico ‘o’ é redundante, dado o relativo já funcionar como objeto do verbo ‘conhecer’ (comparar com *que não o conheci no campo*). No exemplo (29), tratando-se a subordinada de uma oração apositiva de frase, a seleção do

relativo ‘que’ não é adequada, devendo ser substituído por ‘o que’ (comparar com *o que foi brutal*). Como lembra Veloso (2013: 2081), a substituição de ‘o que’ por ‘que’ só “é (marginamente) possível em registos orais (...) e ou em registos escritos não formais”. Por fim, na construção exemplificada em (30), nota-se uma discordância entre o pronome relativo, que recupera o antecedente ‘razões’, no plural, e o verbo da relativa (comparar com *que me fazem ser apologista*). Quanto a casos como o ilustrado em (20), Peres e Móia (1995: 502) afirmam que, apesar de tal tipo de construção já ser normalmente aceite por parte de muitos falantes e de algumas gramáticas, o seu uso deve ser evitado – e, reforço, em contextos escritos formais – pois que, ainda segundo os autores, se trata de uma concordância “anormal no plano sintático e ilógica no plano semântico”. É importante realçar que, como se trata de textos de aprendentes, podem surgir outros desvios além dos referidos, e, por vezes, associados com a construção relativa, mas, por uma questão de foco, deter-me-ei apenas nestes casos.

De acordo com o tipo específico de desvio em relação ao morfema *que*, no registo narrativo, em todos os anos, a estratégia cortadora é a mais comum, sendo representada por 100,00% das ocorrências no 5.º ano; no 7.º, este tipo de desvio corresponde a 50,00%, sendo 12,50% referentes à estratégia de duplo preenchimento e 37,50% a outros desvios; no 10.º ano, a estratégia cortadora equivale a 42,86%, o duplo preenchimento a 28,57% e outros desvios também a 28,57%. No registo argumentativo, o uso inadequado corresponde a valores bem diferentes. Apesar de, no 5.º ano, a estratégia cortadora ser a forma mais recorrente, com 87,50% das ocorrências (enquanto 12,50% são referentes a desvios de outras naturezas), no 7.º, a eliminação de preposição obrigatória passa a ser responsável por apenas 20,00% das ocorrências anómalas (enquanto o duplo preenchimento equivale a 6,67%, sendo a maioria das ocorrências referentes aos outros tipos de desvios, com 73,33% das ocorrências).

As formas de expressão que aqui julgo anómalas já não o são em muitos registos, nomeadamente orais, fazendo parte inclusive da fala de adultos escolarizados, como asseguram Brito e Duarte (2003: 667). Nestes contextos, portanto, já não se pode rotulá-las como marginais. Talvez como consequência desta disseminação de uso, encontram-se ocorrências destes tipos de construção também em vários registos escritos, como, por exemplo, no jornalístico, o que Peres e Móia (1995) ilustram profusamente. Contudo, o facto de sur-

girem nestes meios de comunicação escrita pode colocar em causa o julgamento de inadequação que assumo neste estudo. Ainda assim, o julgamento de inadequação justifica-se pelo facto de o *corpus* investigado ser constituído por textos escritos por aprendentes em contexto formal de escolarização, dos quais se exige, por norma, uma expressão linguística sempre aproximada das expressões cultas da língua, francamente descritas (e prescritas) nas gramáticas e nos manuais didáticos de referência para o ensino de português. Subscribo ao que afirmam Peres e Mória (1995: 41), segundos os quais a anomalia resulta “da falta de familiaridade com os monumentos escritos da língua ou da ausência de um distanciamento em relação a ela que permita a compreensão da sua orgânica e das imensas possibilidades que oferece.” Ambas as causas constituem-se como preocupação dos princípios do ensino de línguas em contexto escolar.

6.8. Orações adverbiais

Como comprovado pelo estudo de correlação e ilustrado pelo Gráfico 24, abaixo, a frequência de orações adverbiais, no *corpus*, diminui em ambos os registos à medida que os alunos avançam nos anos de escolaridade, caracterizando-se esta diminuição por ser não gradual do ano inicial até ao ano final, já que há uma acentuada diminuição após o 7.º ano. Em termos percentuais, e em complementação aos valores de rácio já apresentados, as orações adverbiais representam no 5.º ano dos textos narrativos 45,53% do total de orações subordinadas; no 7.º, 43,51%; e, no 10.º, 33,20%. No registo argumentativo, equivalem no 5.º ano a 36,93%; no 7.º, a 36,26%; e, no 10.º, a 25,71%.

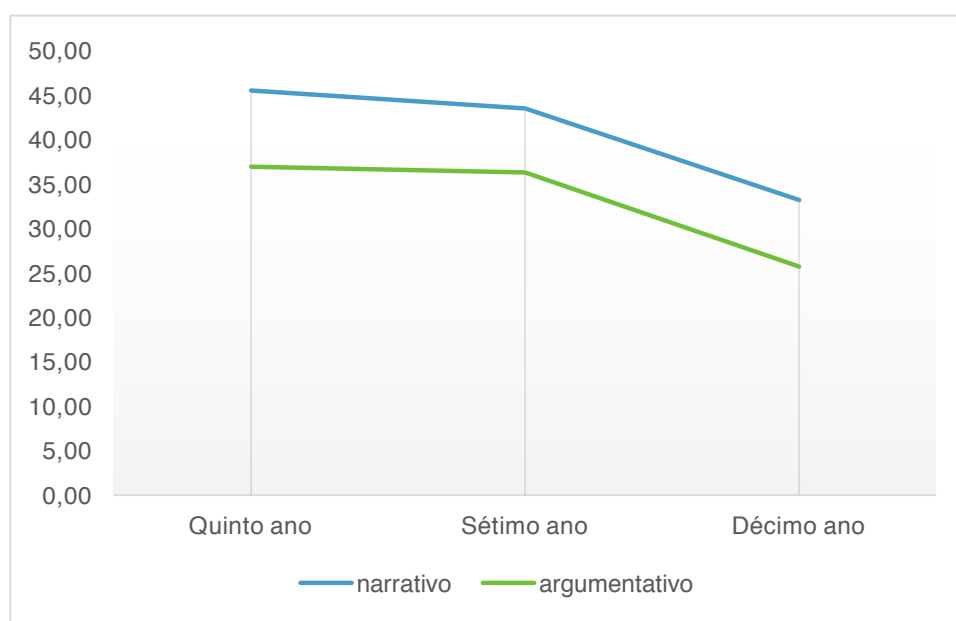


Gráfico 24: Distribuição dos valores médios de ocorrências, em percentual, de orações adverbiais nos registos narrativo e argumentativo ao longo dos anos escolares.

Mesmo com a diminuição na frequência de uso, as adverbiais são o tipo de subordinada prevalente no 5.º e no 7.º ano, cedendo espaço a outros tipos de subordinadas apenas no 10.º ano, o que reflete, pelo menos aproximadamente, a afirmação de Perera (1984, cf. Allison *et al.* 2002: 104) de que as adverbiais ocorrem extensivamente em textos escritos por crianças e adolescentes entre os sete e os dezassete anos. É a ocorrência expressiva de orações adverbiais no *corpus*, nomeadamente no registo narrativo, que motiva este estudo de caso.

As orações adverbiais não são selecionadas pela oração a que se subordinam, antes ocupando uma posição de adjunção. Embora haja um conjunto de propriedades de natureza sintática que as possam caracterizar, como, por exemplo, a possibilidade de estarem antepostas à oração subordinante, ou mesmo de ocorrerem entre o sujeito e o predicado, propriedade que as distingue das outras subordinadas, são tipicamente classificadas pelos valores semânticos que veiculam em relação ao domínio proposicional. Lobo (2013: 1982 e seguintes) refere, para além dos tradicionais subtipos de oração subordinada adverbial (condicionais, causais, finais, concessivas e temporais), outros, que não costumam figurar em gramáticas prescritivas ou descritivas. São elas as orações de circunstância negativa, de comentário, substitutivas e acrescentativas. Há ainda um subtipo de oração não referida nas gramáticas de referência, mas com ocorrência no *corpus*. A função desta oração, tipicamente iniciada por *como*, consiste em exemplificar uma informação estabelecida na oração principal. Designarei este tipo de oração por exemplificativa e, por simplificação, incluirei este tipo de orações na classe das orações adverbiais. (cf. Anexo 9.7).

Face a esta gama extensa de significados que as orações adverbiais codificam, é importante aferir se a diminuição do uso ocorre na totalidade dos subtipos de adverbiais ou somente em alguns, considerando-se adicionalmente o registo como um fator. O Gráfico 25 e o Gráfico 26, abaixo, mostram como cada subtipo de oração subordinada, a partir da classificação adotada, que inclui classes propostas por Lobo (2013), se comporta ao longo dos anos escolares estudados, respetivamente no registo narrativo e no registo argumentativo:

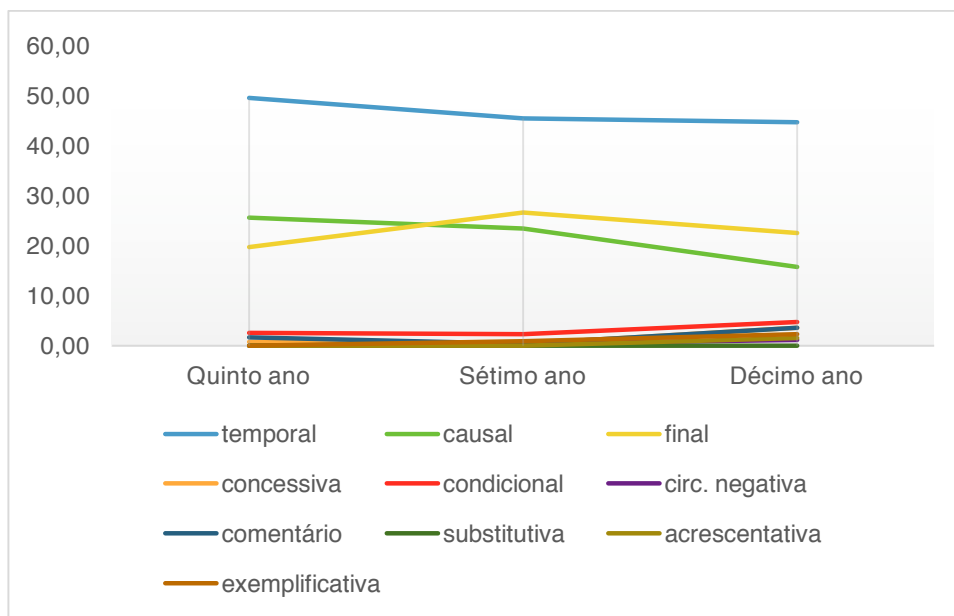


Gráfico 25: Distribuição dos valores médios de ocorrências, em percentual, dos tipos de orações subordinadas adverbiais no registo narrativo ao longo dos anos escolares.

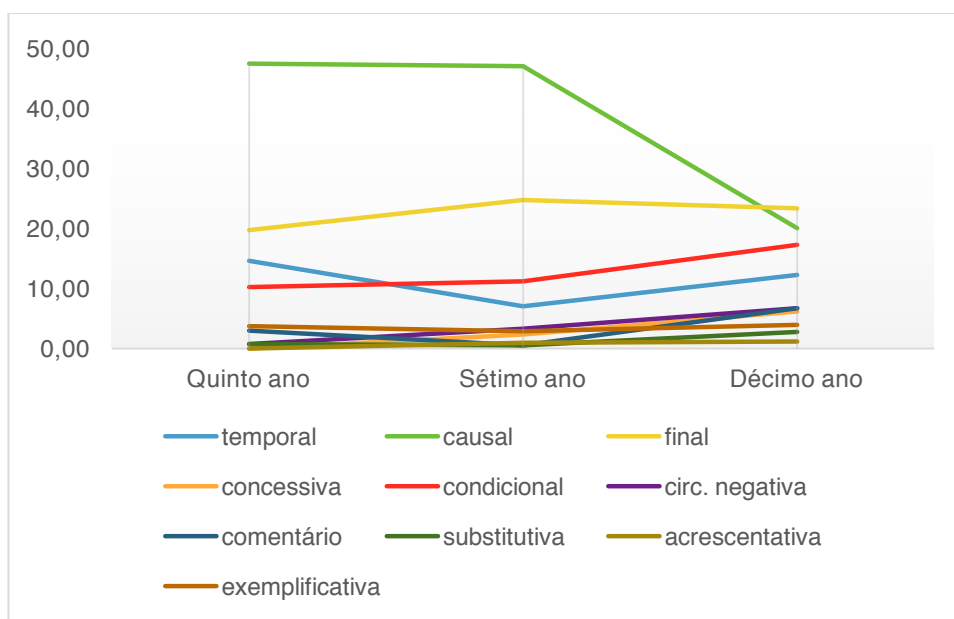


Gráfico 26: Distribuição dos valores médios de ocorrências, em percentual, dos tipos de orações subordinadas adverbiais no registo argumentativo ao longo dos anos escolares.

No registo narrativo, como se pode ver no Gráfico 25, é expressiva (e até esperada, por se tratar deste registo) a utilização de orações adverbiais temporais, representadas, no 5.º ano, por 49,57% do total de ocorrências; no 7.º, por 45,41%; e, no 10.º, por 44,66%. São também elevados os valores percentuais de orações causais e finais. As primeiras, no 5.º

ano, representam 25,64% do total de ocorrências: no 7.º, 23,39%; e, no 10.º, 15,81%. As últimas equivalem, no 5.º ano, a 19,66%; no 7.º, a 26,61%; e, no 10.º, a 22,53%. Estes valores percentuais das orações causais e finais, que, somados, ultrapassam 45% do total de ocorrências, parecem não se enquadrar no que afirmam Gonçalves *et al.* (2011: 46-7), segundo as quais as orações adverbiais, excetuando-se as temporais, são de uso restrito no quarto ano de escolaridade. Vale lembrar – como explicado na Secção 3.1 – que os textos do 5.º ano do *corpus* foram coletados no início do ano letivo, a fim de que representassem padrões de ocorrência relativos aos estádios de aprendizagem do primeiro ciclo, sem grandes interferências do novo ano/ciclo escolar. O Gráfico 25 mostra ainda que os outros subtipos de adverbiais têm um emprego, em termos quantitativos, muito semelhante no 5.º e no 7.º ano, sendo a soma das suas ocorrências igual a 5,13% no ano inicial e 4,59% no ano medial. No 10.º ano, estes percentuais somados apresentam um aumento significativo, passando a 17,00% do total de orações adverbiais.

No registo argumentativo, a distribuição ao longo dos anos de escolaridade mostra-se menos bem definida, como se vê nos textos de registo narrativo. O Gráfico 26 mostra que as orações causais são o principal tipo de orações subordinadas adverbiais presentes no *corpus*, nomeadamente no 5.º e no 7.º anos, passando, no 10.º ano, as orações finais a ocupar a posição dianteira entre as orações adverbiais, embora haja um crescimento dos outros tipos, constituindo-se, neste ano escolar, um cenário de distribuição mais fragmentado, o que sugere, por consequência, um uso mais equânime das adverbiais. Em termos percentuais, no 5.º ano, do total deste tipo de orações, as causais preenchem 47,45%; no 7.º, 46,98%; e, no 10.º, 20%. As orações finais, também bastante representadas, correspondem, no 5.º ano, a 19,71% do total; no 7.º, a 24,65%; e, no 10.º, a 23,33%. É de destacar ainda o emprego das orações de valor temporal neste registo, correspondendo, no 5.º ano, a 14,60% do total de orações adverbiais; no 7.º, a 6,98%; e, no 10.º, a 12,22%.

Os estudos dos casos de uso que a seguir se apresentam estão relacionados com as orações de mais expressiva ocorrência no *corpus*, ou seja, as orações temporais no registo narrativo e as causais no registo argumentativo.

6.8.1. Localização temporal entre situações e posição das orações temporais

Este breve caso de uso é motivado inicialmente pela seguinte afirmação de Diessel (2008: 465): “... the linear ordering of main and subordinate clauses mirrors the sequential

ordering of the events they describe.” Para este autor (a par de Chafe 1984, Thompson 1985 ou Ford 1993), na organização do discurso, a posição da oração adverbial em relação à principal regula-se por fatores de natureza pragmática, entre outros: a oração adverbial em posição final tende a apresentar uma informação nova; já em posição inicial serve para organizar o fluxo informacional (Diessel 2008: 467). Esta convicção fundamenta-se num conjunto de estudos baseados em *corpora* de textos falados e escritos em diferentes línguas (Altenberg 1984, Ramsay 1987; Ford 1993; Diessel 2001, 2005) em que se analisam orações adverbiais condicionais, causais, finais, concessivas e temporais em diferentes posições (inicial, medial ou final) relativamente à principal. Brito (2003: 698) faz o mesmo tipo de afirmação para o português, enquanto Lobo (2013: 1991) refere que as orações adverbiais ocorrem, “na maioria dos casos”, em posição final, que é, portanto, a posição não-marcada. Repara esta autora (2013: 2037) que à posição das adverbiais “estão associados valores discursivos distintos”, delimitando-se uma perspetiva semelhante à defendida por Diessel.

Sem subscrever aos pressupostos teóricos subjacentes à afirmação de Diessel e focando-me apenas nas orações temporais, interessa-me aqui, de acordo com o espírito desta dissertação, descrever comparativamente os dados relativos à posição desta oração adverbial (finita) em relação à sua subordinante (simplificadamente, posição inicial, medial ou final) e os dados relativos à localização do intervalo de tempo em que ocorre a situação expressa na oração adverbial em relação à do intervalo em que ocorre a situação expressa na sua subordinante (simplificadamente, uma relação de anterioridade, posterioridade ou sobreposição). Dito doutro modo, interessa aqui saber se, como asserem Gonçalves *et al.* (2008: 46, remetendo a Sim-Sim 1998), existe alguma tendência, nos textos escritos narrativos do *corpus*, de “considerar que o que surge (...) em primeiro lugar é o que acontece primeiro”.

A localização temporal consiste no posicionamento de situações⁵⁴ no eixo do tempo (Móia 2003: 108). Entre as várias possibilidades, encontra-se a de localizar uma situação em relação a uma outra, para o que contribuem as orações temporais, definindo-se, portanto, uma relação entre o tempo da situação relevante identificada por uma oração e o tempo da

⁵⁴ Opto pela expressão “situações” como um sinónimo de eventualidades, estados-de-coisas, eventos, etc.

situação identificada pela oração a que ela se subordina (Quirk *et al.* 1985: 1080). A relação temporal entre estas duas situações pode ser de sobreposição ou de sucessão⁵⁵, isto é, o intervalo de tempo em que se localiza a situação relevante descrita pela oração subordinada pode ser anterior àquele em que se localiza a situação descrita pela oração principal (cf. (31), (32), (33) e (34)) ou posterior (cf. (35) e (36)⁵⁶), ou ainda sobreposto (cf. (37), (38) e (39)). Vejam-se os exemplos, com as orações subordinadas demarcadas por parênteses retos e os conectores temporais em itálico:

31. ... não vou dizer que ela ficou bem, mas sempre ficou menos triste [*depois de* conversar-mos.] (afcf_1_10h_atd)
32. ... [*quando* vimos o caminho lá ao longe] apercebemo-nos que para chegar lá tínhamos que passar por uma estrada em obras... (mit_1_10a_fdl)
33. ... [*ao* acabar de comer,] conseguimos encontrar as nossas malas... (dmc_1_7b_mda)
34. ... [*quando* eu e a minha melhor amiga chegamos,] eles nos disseram ‘olá’... (beb_1_5h_mda)
35. [*Antes de* o largar] eu e a minha prima tiramos-lhe a coleira. (mf3_1_7a_atd)
36. [*Quando* construíram a ponte,] contrataram arquitetos famosos.
37. Os adultos iam de manhã à escola de Golfe Executivo, [*enquanto* nós íamos relaxar apanhando banhos de sol e indo à piscina do hotel.] (pp_1_7a_atd)
38. [*Quando* eu era pequeno,] adorava a minha mãe. (gaco_1_5h_mda)
39. ... [*quando* já era de noite,] fomos dormir... (mmb_1_5h_mda)

⁵⁵ “Sobreposição” e “sucessão” são rótulos genéricos, havendo, sob cada um deles, definições de localização mais finas. Para aprofundamentos, veja-se Móia (2000) ou Móia e Alves (2013).

⁵⁶ Exemplo extraído de Cunha (2000: 311), traduzido de Moens e Steedman (1987).

Os itens conetores, como se depreende dos exemplos acima, podem contribuir para a explicitação da localização temporal que se estabelece entre as situações descritas na oração subordinada e na subordinante. Alguns têm significados temporais intrínsecos, quer de anterioridade ou posterioridade (‘depois de’, ‘antes de’), quer de sobreposição (‘enquanto’). Com outros, porém, a relação temporal que a situação descrita pela oração subordinada estabelece com a que é descrita pela principal depende de diversos outros fatores, tais como a classe aspetual dos predicados (cf. (26)), os tempos gramaticais (cf. (25)) e o conhecimento de mundo (cf. (23)) (Lobo 2013: 2001). É esta a característica, por exemplo, do conector ‘quando’, que, segundo Oliveira (2003: 177-8) é “um localizador mais ou menos neutro quanto à determinação da ordem relativa entre as situações envolvidas, sendo a sua principal função a de saturar, com a eventualidade que lhe está associada, as possibilidades de localização temporal da oração principal, sem impor restrições adicionais em termos de relação temporal entre as situações”. Pelo caráter polifuncional na codificação da localização temporal e pela ocorrência expressiva no *corpus*, centra-se esta discussão neste conector, sobre o qual é necessária ainda uma nota de esclarecimento.

Antes de serem apresentados os dados relativos ao uso do conector ‘quando’, vale a pena reparar que a inclusão no grupo das orações adverbiais temporais das orações subordinadas encabeçadas por este conector não é livre de discussão. Hoje convivem duas classificações que a literatura linguística oferece como possíveis. Ilustra esta convivência a Gramática do Português (2013). Por um lado, Lobo (2013: 2003), no capítulo sobre as orações adverbiais, toma as orações aqui em causa como adverbiais temporais, uma classificação de base semântica já sacralizada nas gramáticas tradicionais. Por outro lado, Veloso (2013: 2107), no capítulo sobre as relativas, ecoa as proposições de Mória (1992, 2001), que, fundamentado em propriedades sintáticas do português, opta por considerá-las como orações relativas. As duas classificações são sobejamente fundamentadas, mas, do ponto de vista da investigação, a escolha por uma ou outra, obviamente, tem custos. Ao assumir-se que são orações adverbiais, perde-se a possibilidade de juntar este tipo de orações a um conjunto de orações relativas que são suas paráfrases francas, como se vê em (40), repetição de (32), e (41):

40. ... [*quando* vimos o caminho lá ao longe] apercebemo-nos que para chegar lá tínhamos que passar por uma estrada em obras... (mit_1_10a_fdl).

41. [No momento em que vimos o caminho lá ao longe] apercebemo-nos que para chegar lá tínhamos que passar por uma estrada em obras.

Optar por incluí-las no conjunto das relativas, não obstante satisfazer a coerência sintática, pode impedir que o estatuto semântico das construções de valor temporal iniciadas por ‘quando’ tenha implicações na interpretação do discurso produzido nos textos escolares aqui em apreço, e em particular nos textos de modalidade narrativa, que culturalmente refletem as diretrizes de programas curriculares, materiais didáticos e mesmo de práticas de sala de aula do ensino de português. É por esta razão que, neste estudo, inserido no âmbito da linguística aplicada à educação, se segue a classificação tradicional. São exceção os casos em que o item *quando* tem estatuto de argumento da oração matriz. Nestes casos, assumo a oração subordinada como relativa, como se vê segmentado em (42):

42. Outro momento muito aplaudido foi [quando um congressista de Baucau se lembrou de destacar algumas das marcas dos sistemas democráticos, ou seja, a separação de poderes]. (CRPC-J40716)

Feita a observação, mostra-se, no Gráfico 27, a diversidade de uso dos conectores das orações temporais identificada nos textos narrativos do *corpus*:

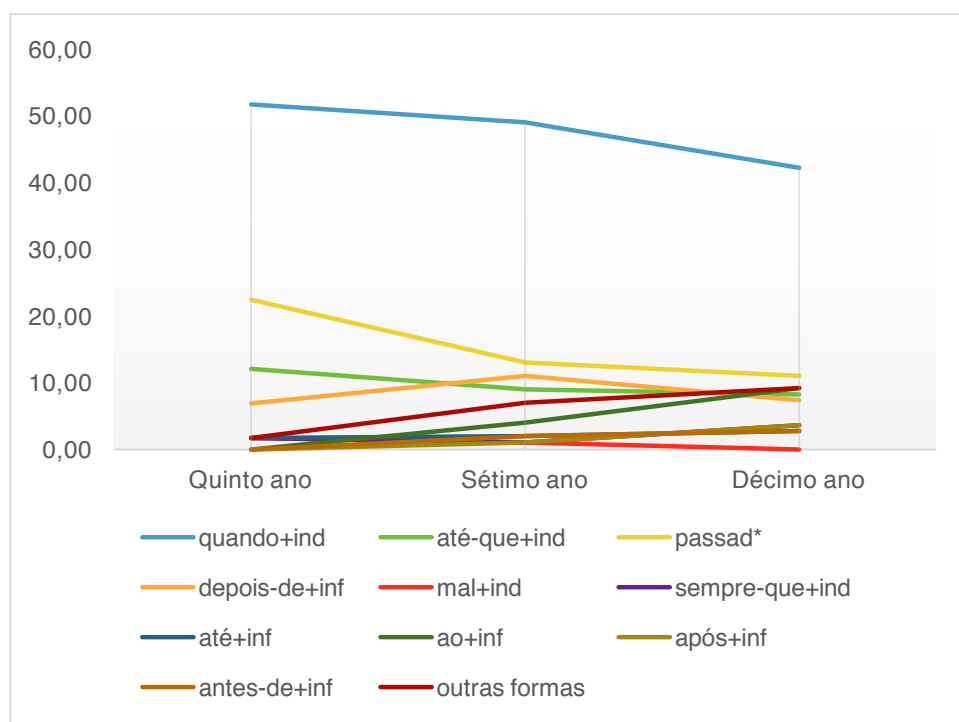


Gráfico 27: Distribuição dos valores médios de ocorrências, em percentual, de conectores empregues nas orações temporais no registo narrativo ao longo dos anos escolares.

A dinâmica da distribuição dos conectores de orações temporais nos textos narrativos sugere que a seleção destes itens, do 5.º ao 10.º ano, descentraliza-se. Enquanto no 5.º ano, as formas ‘quando’, ‘até que’ e ‘depois de’ e as orações participiais (concretamente as iniciadas com a forma nominal do verbo ‘passar’) correspondem a 93,10% do total de ocorrências de conectores, no 7.º passam a representar 82,00% e, no 10.º, 68,81%, indicando, portanto, a seleção de outras formas possíveis. Embora as ocorrências do item ‘quando’ diminuam ao longo da progressão escolar – o que parece estar em consonância com a correlação negativa entre as orações adverbiais e a progressão escolar (cf. Secção 5.5) e, em particular, com a diminuição do emprego de orações temporais anteriormente descrito (cf. Gráfico 25) –, este item constitui a principal escolha dos escritores para introduzir as orações de valor temporal. De acordo com Gonçalves *et al.* (2011: 46-7), as orações temporais são já de ocorrência frequente em textos narrativos escritos por alunos portugueses no quarto ano de escolaridade, o que se confirma, por exemplo, em Almeida (2013: 75), com base num conjunto de textos narrativos escritos por crianças do quarto e do sexto ano de escolaridade (e também na aplicação de testes experimentais).

A partir dos três modos fundamentais de localização temporal entre situações (anterioridade, posterioridade e sobreposição) e tendo em conta o item ‘quando’, identificou-se que, nos textos do 5.º ano do *corpus* narrativo, 61,29% do total de orações temporais expressam uma relação de anterioridade (em relação ao intervalo de tempo da situação relevante descrita pela oração subordinante); no 7.º, este valor passa a 88,89%; e no 10.º, a 76,92%. Quanto às orações temporais que expressam uma relação de posterioridade, há ocorrências apenas no 5.º ano, com 3,23%, e no 7.º ano, com 6,67%. Por fim, considerando-se a localização temporal de sobreposição, têm-se 35,48%, no 5.º ano; 4,44%, no 7.º ano; e 23,08%, no 10.º ano. Excluem-se destas ocorrências as orações temporais a que Lobo (2013: 1983) chama de narrativas. Embora a situação relevante descrita nesta subordinada esteja localizada temporalmente dentro do intervalo de tempo da situação descrita na subordinante, a sua posição é obrigatoriamente sempre à direita desta, já que representa um clímax dramático narrativo (Quirk *et al* 1985: 1084), como se exemplifica em (43), abaixo:

43. Eu e meu o meu melhor amigo estávamos no autocarro [quando à nossa frente se senta uma mulher] ... (bfbc_1_7b_mda)

Quanto às possibilidades de posicionamento da oração adjunta temporal (inicial, medial ou final) em relação à oração subordinante, identificou-se que: no 5.º ano, do total de orações temporais iniciadas pelo conector ‘quando’, 80,65% estão em posição inicial; no 7.º, 97,78%; e, no 10.º, 82,05%. As orações em posição final representam, no 5.º ano, 16,13%; no 7.º, 2,22%; e, no 10.º, 15,38% do total. Por fim, no 5.º ano, apenas 3,23% deste tipo de orações ocorre em posição medial; no 7.º, não houve ocorrências; e, no 10.º, 2,56%.

De modo a enriquecer a discussão proposta neste breve caso de uso, em que se pretende, com base no *corpus*, descrever os dados da relação entre, por um lado, a localização temporal (anterioridade, posterioridade ou sobreposição) expressa pela oração adverbial temporal e, por outro lado, a sua posição (inicial, medial ou final), primeiramente, ilustra-se, no Gráfico 28, o movimento de distribuição dos tipos de relação identificadas ao longo dos anos escolares estudados, para, de seguida, se apresentarem os valores percentuais:

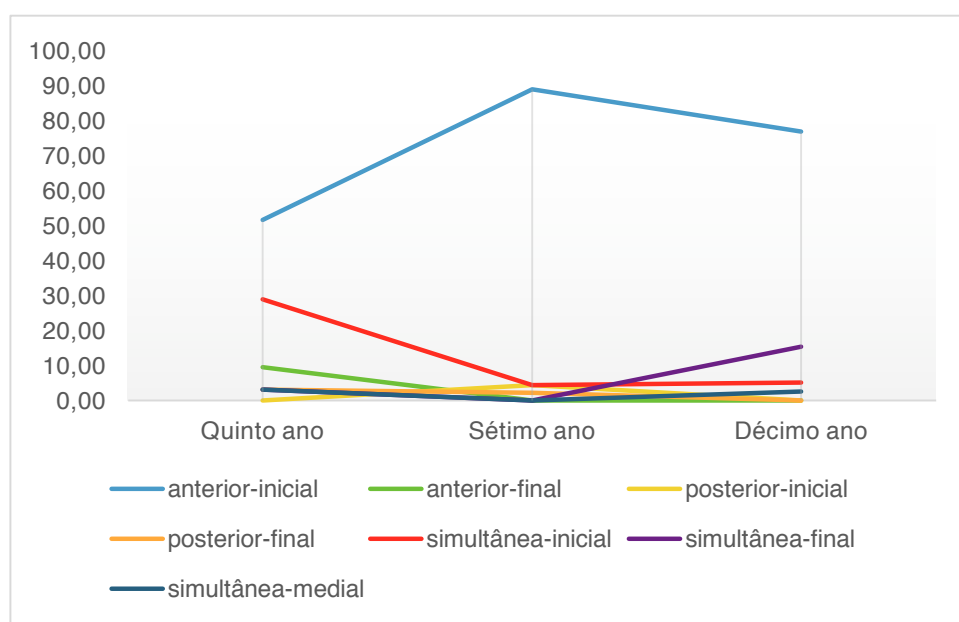


Gráfico 28: Distribuição dos valores médios de ocorrências, em percentual, dos tipos de relação entre a localização temporal da oração temporal (anterioridade, posterioridade e sobreposição) e a sua posição (inicial, medial e final) no registo narrativo ao longo dos anos escolares.

No gráfico, pode ver-se que a principal relação expressa nos textos narrativos dos alunos é a relação anterior-inicial, com um crescimento não-linear de uso do 5.º ao 10.º ano,

havendo um pico de ocorrência bastante expressivo nos textos do 7.º ano, em comparação com os outros anos. Em termos percentuais, no 5.º ano, 51,61% das relações são a do tipo anterior-inicial; no 7.º, o percentual é de 88,89%; e, no 10.º, de 76,92. Com a mesma localização temporal, mas em posição final (relação anterior-final), só há ocorrências no 5.º ano, representadas por 9,68% das ocorrências totais. Não há, em nenhum dos anos escolares estudados, ocorrências da relação anterior-medial. A relação do tipo posterior-inicial foi encontrada somente no 7.º ano, correspondendo a 4,44% do total de orações temporais iniciadas por ‘quando’. A relação posterior-final também tem ocorrência restrita, somente aparecendo no 5.º ano, com 3,23%, e, no 7.º ano, com 2,22%. Não há, em nenhum dos anos escolares, relações do tipo posterior-medial. A relação simultânea-inicial representa-se, no 5.º ano, por expressivos 29,03% do total de ocorrências; no 7.º ano, decai para 4,44%; e, no 10.º ano, atinge 5,13%. Tanto a relação simultânea-final como a simultânea-medial só acontecem no 5.º ano, em que cada uma das relações equivale a 3,23% do total, e no 10.º ano, em que a relação simultânea-final responde por 15,38% do total de ocorrências, enquanto a relação simultânea-medial responde por 2,56%.

Em síntese, pode-se dizer que os valores percentuais apresentados no Gráfico 28 convergem para a conclusão de que há uma tendência de se realizar a oração temporal que expressa anterioridade (em relação à situação descrita pela oração subordinante) numa posição inicial (relação anterior-inicial), embora seja esta tendência cada vez menos expressiva ao longo da progressão escolar. Não se pode, contudo, concluir que haja uma sistematicidade de ocorrências entre a expressão da localização temporal e a posição da oração temporal (ou, *grosso modo*, entre a questão semântica da localização temporal e a questão sintática da posição da oração temporal na estrutura frásica). O segundo valor mais substancial no 5.º ano é o da relação simultâneo-inicial e, no 10.º, da relação simultâneo-final. Não se podendo indicar as causas da preferência pela relação anterior-inicial (e não sendo esse o objetivo deste estudo), a consideração de que a posição inicial da oração temporal encabeçada por *quando*, independentemente da localização temporal, é expressivamente a principal escolha em todos os anos escolares, como apresentado anteriormente, permite apenas concluir que, nos textos narrativos investigados, se estabelece a prática linguística de se antepor a oração temporal à sua subordinante, prática que se vê em estudos de referência (Cunha 2000; Brito 2003; Gonçalves *et al.* 2011; Lobo 2013), em que a maioria dos exemplos de orações temporais, apesar de constantemente ser referida a sua propriedade de ser

móvel, estão em posição inicial. Portanto, não será descabido perguntar se não é nesta prática que reside a justificação para a dificuldade, referida por Pyykkönen e Järvikivi (2012: 521) e citada em Almeida (2013: 34), que crianças entre sete e doze têm de compreender sequências oracionais quando a localização das situações não é coincidente com a posição na linearidade frásica.

6.8.2. Diversidade de uso dos conectores das orações causais

Como é visível no Gráfico 26, as orações causais constituem a principal forma de oração adverbial usada nos textos do registo argumentativo do 5.º e do 7.º ano. No 10.º ano, como um potencial reflexo da diminuição global de peso do grupo das orações adverbiais (cf. Gráfico 8), também as causais têm o seu emprego diminuído, embora continuem a aparecer, juntamente com as finais, como orações de ocorrência superior aos outros subtipos. É esta expressiva ocorrência das orações causais (finitas e não-finitas) que justifica descrever a variação de uso dos conectores que as iniciam.

Antes de apresentar os dados relativos a estas ocorrências, é importante salientar que, nesta dissertação, se fundamenta a classificação das orações subordinadas adverbiais causais, do mesmo modo como se faz com todas as outras adverbiais, primeiramente nos significados primários – e mais evidentes – do nexos entre a subordinante e a subordinada, em consonância com o proposto por muitos estudos de referência, como, por exemplo, Quirk *et al.* (1985), Brito (2003), Cunha e Cintra (2008) ou Bechara (2009), e também com o proposto pelo Dicionário Terminológico (DT), segundo o qual as orações subordinadas causais exprimem “a razão, o motivo (a causa) do evento descrito na subordinante” ou apresentam “uma justificação para o que é expresso na subordinante”.

É, portanto, uma classificação de ampla abrangência, que considera não somente as orações de valor causal prototípico, “em que o nexos de causa-efeito se estabelece entre situações do mundo linguisticamente representadas,” como define Lopes (2012: 454), ou em que se verifica “uma ordenação temporal entre os eventos da subordinada e da matriz”, nas palavras de Lobo (2003: 69) (cf. 44)), mas também as orações de valor explicativo (cf. (45)), casos em que, como diz Lopes (2012: 453), e na mesma linha de Sweetser (1999), Peres (1997) ou Peres e Mascarenhas (2006), “o nexos causal não é estabelecido com a predicação expressa na (...) oração principal, mas sim com a enunciação dessa mesma predica-

ção...”, em que está envolvido o mundo do raciocínio do falante, ancorado nos seus conhecimentos e/ou crenças. Podem atuar como conetores, em ambos os contextos, os itens ‘porque’, ‘pois’, ‘como’, ‘visto que’, ‘já que’, ‘uma vez que’ e ainda a construção ‘por’ + INF. É importante reiterar que o facto de estes conetores serem aqui classificados como subordinativos não afeta as potenciais leituras distintas que se possam extrair das orações que encabeçam, ou seja, entende-se que o ‘porque’ ou o ‘pois’, por exemplo, podem funcionar tanto como conetores de orações causais prototípicas (cf. 46)), como de explicação, justificação ou de dedução (cf. (47)).

44. Então cada vez que chegava a hora de ir para a praia nós escondíamo-nos sempre, [*porque* tínhamos muito medo que o tubarão-martelo estivesse naquela praia]. (basb_1_5c_mda)
45. No Google+, nós devemos ter cuidado com o que vemos [*porque* pode ter vírus e *porque* pode não ser coisas próprias para a nossa idade]. (mcolfs_2_5h_mda)
46. Certo dia de verão tinha eu acabado de acordar, liga-me o Francisco, estava todo empolgado [*pois* tinha lhe sido dito que uma discoteca de Lisboa iria abrir pela primeira vez naquela noite] ... (pmls_1_10h_atd)
47. ... é bom termos as redes sociais [*porque* podemos conhecer melhor os nossos conhecidos]. (bg_2_7a_hnn)

Como é óbvio, não se toma como incontroversa a inclusão das orações de valor explicativo no grupo das orações subordinadas que aqui se faz, em particular com relação às orações iniciadas por ‘porque’, ‘pois’ e ‘que’. Lobo (2001: 294), por exemplo, refere-se a estas orações como construções de coordenação, propondo, a partir de Quirk *et al.* (1985), Piot (1988), López Garcia (1999), um conjunto de testes sintáticos para a identificação de estruturas oracionais coordenadas, de que se pode citar a impossibilidade de este tipo de oração ocupar a posição inicial⁵⁷. Quanto ao critério da anteposição da oração explicativa, Brito (2003: 713) admite que “há alguma hesitação nos juízos de gramaticalidade” da frase

⁵⁷ Outros testes referidos por Lobo (2001: 294) são a impossibilidade de a oração ser coordenada, a impossibilidade de os clíticos ocorrerem em próclise e a impossibilidade de a conjunção ligar constituintes não-frásicos.

resultante, concluindo, no caso do conector ‘porque’, que há alguns comportamentos que permitem classificá-lo como subordinativo, do mesmo modo como faz Matos (2004). Os tradicionais critérios sintáticos utilizados para distinguir a coordenação da subordinação podem falhar em particularizar alguns tipos de orações, para além de que a tentativa da distinção entre coordenação e subordinação é ela própria, por vezes, geradora de forte controvérsia, como observa, e.g., Lopes (2012: 453).

A dificuldade que resulta da tradicional (e restritiva) distinção binária entre coordenação e subordinação tem como consequência propostas de classificação alternativas, como as de Foley e Van Valin (1984) (coordenação, subordinação e co-subordinação), as de Huddleston e Pullum (2007) (coordenação, subordinação e suplementação) ou ainda as de Peres e Mascarenhas 2006 (coordenação, subordinação, anexação e suplementação), entre outras. A existência destas propostas de classificação contraria o mito de que a distinção entre a coordenação e a subordinação é um caso encerrado. A postura que se assume aqui, em que se considera a causa num sentido lato, é apenas pela necessidade de um tratamento homogéneo, já que é um estudo baseado num *corpus* de aprendentes (e relativamente extenso), em que algumas das ocorrências nos textos argumentativos tanto permitem que se considere que se está perante uma oração de enunciação, como de uma oração de enunciado. Deste modo, evitam-se, sempre que possível, avaliações cuja subjetividade conduza a incoerências na anotação das ocorrências linguísticas.

Não é parte dos objetivos desta dissertação discutir o estatuto sintático (se coordenado ou subordinado) das orações explicativas, tampouco o é apresentar uma nova proposta de classificação. Para mais sobre este debate para o português, remeto para os autores citados, e também Lobo (2001), Peres e Móia (2003), Colaço (2005) ou Lopes (2004); para outras línguas, Quirk *et al.* (1985), Renzi e Salvi (1991) ou Galán Rodríguez (1999), por exemplo.

Feitas estas observações sobre a classificação das orações causais aqui adotada, apresenta-se de seguida, no Gráfico 27, a distribuição dos usos dos conectores causais (entendidos no sentido lato, que abrange os valores explicativos) nos textos argumentativos ao longo dos anos escolares:

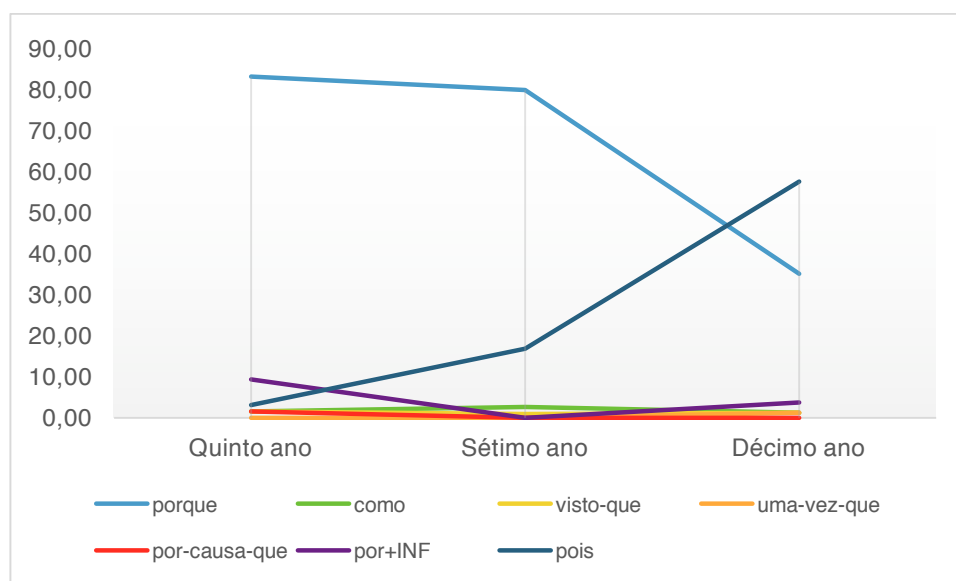


Gráfico 29: Distribuição dos valores médios de ocorrências, em percentual, dos tipos de conectores empregues nas orações causais no registo argumentativo ao longo dos anos escolares.

Do Gráfico 29, destaca-se a seleção dos conectores ‘porque’ e ‘pois’ para encabeçar as orações subordinadas causais, já que há um movimento de inversão no uso ao longo da progressão nos anos escolares. O valor percentual médio de ocorrência de ‘porque’, que é acentuada no ano inicial, correspondendo a 83,08% do total de conectores causais empregues, decaindo sensivelmente até ao ano final, equivalendo, no 7.º ano a 79,83% e, no 10.º, a 35,00%. O conector ‘pois’, que, num processo contrário, tem uma ocorrência minúscula no ano inicial, representando 3,08% do total de conectores causais, eleva-se significativamente até ao ano final, correspondendo, no 7.º ano, a 16,81% e, no 10.º, a 57,50%. Todos os outros conectores causais têm ocorrências inexpressivas se comparadas com a utilização de ‘porque’ e ‘pois’, excetuando-se a ocorrência da construção *por+INF*, que, embora seja a segunda ocorrência mais expressiva nos textos argumentativos do 5.º ano (9,23%), não ocorre nos textos do 7.º ano e reduz-se a 3,75% no 10.º ano. Assim, com base nas ocorrências de ‘porque’ e ‘pois’, pode concluir-se que a maioria das orações causais se realiza com verbos finitos, correspondendo no 5.º ano a 86,15% do total de ocorrências, no 7.º a 96,64% e no 10.º ano a 92,50%.

Esta situação identificada no *corpus*, isto é, de pouca variação na seleção do conector de subordinação causal, tendência constante ao longo dos três anos escolares, parece não refletir um dos resultados que se espera ver na produção linguística dos alunos de acordo

com o Programa de Português (Portugal 2009: 117), segundo o qual, ao fim do terceiro ciclo, os alunos devem ser capazes de “produzir textos em português padrão, recorrendo a vocabulário diversificado e a estruturas gramaticais com complexidade sintáctica...”. Saliente-se ainda que, no mesmo programa, o conteúdo relativo às conjunções causais constitui parte do eixo relativo ao conhecimento explícito da língua ainda no primeiro ciclo, sendo retomado, relativamente ao mesmo eixo, no segundo ciclo, e não apenas como parte do processo de ensino das conjunções do português, mas também como parte dos processos sintáticos de articulação de orações.

Dado este cenário, poder-se-ia questionar se os alunos estudados desconhecem outros conectores causais, para além do *porque*, ou se se trata apenas de seleção, de escolha (consciente ou inconsciente). Tentar responder a esta questão é um afastamento da essência deste estudo, que consiste na descrição de frequências, e não na atribuição de causas, mas a apresentação da utilização que os alunos fazem do mesmo tipo de conectores no registo narrativo, em que a ocorrência, em termos numéricos, também é expressiva, pode mostrar-se interessante. O Gráfico 30 apresenta tais dados:

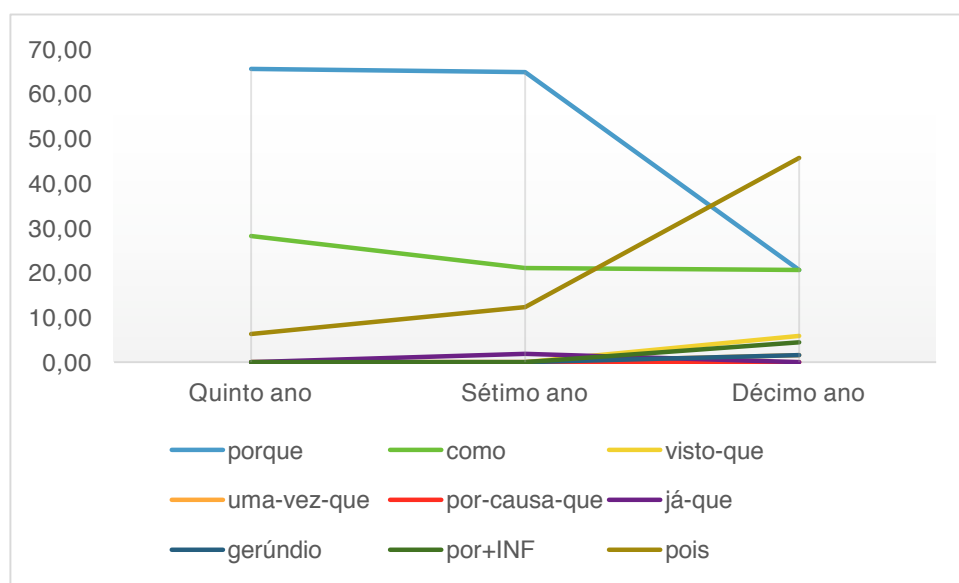


Gráfico 30: Distribuição dos valores médios de ocorrências, em percentual, dos tipos de conectores empregues nas orações causais no registo narrativo ao longo dos anos escolares.

A dinâmica da distribuição que se vê no registo narrativo assemelha-se à apresentada para o registo argumentativo, em que há o abandono gradual do conector ‘porque’, que representa, no 5.º ano, 65,63% do total de ocorrências, no 7.º, 64,91% e, no 10.º, 20,59%, e

da elevação expressiva do conector ‘pois’, que, no 5.º ano, equivale a 6,25% do total, no 7.º a 12,28% e no 10.º, a 45,59%. No entanto, outra forma também tem realização expressiva, nomeadamente o item ‘como’. Este item, no 5.º ano, equivale a 28,13% do total, no 7.º, a 21,05% e no 10.º a 20,59%. Decorre do Gráfico 30 que, nos textos narrativos, os outros conectores, apesar de aparente elevação no uso, continuam a ser de pouco acionados na escrita dos alunos.

Tendo em conta a tarefa de escrita que serve de estímulo para os textos argumentativos⁵⁸, em que se pede a opinião dos alunos sobre a importância das redes sociais, a utilização quase isolada dos conectores ‘porque’ e ‘pois’ (em alternância, a depender do ano escolar) parece refletir a sua função discursiva, enquanto conectores de uma oração causal de enunciação. Como refere Lopes (2012: 456), este tipo de orações “justifica a asserção inicial, justamente porque o seu conteúdo proposicional é apresentado como argumento válido para sustentar a conclusão inferida pelo locutor.” Entende-se, aqui, que a asserção original está presente na tarefa, sendo assinalada pela presença do verbo ‘achar’. É sempre possível, ainda em concordância com Lopes, acionar, em potenciais paráfrases (cf. (48), paráfrase de (47), e (49), paráfrase de (45)), o verbo epistémico que fundamenta o caráter de enunciação, assemelhando-se estas paráfrases a muitas das ocorrências naturais (cf. (50)):

48. *Acho que é bom termos as redes sociais [porque podemos conhecer melhor os nossos conhecidos].*
49. *Acho que no Google+, nós devemos ter cuidado com o que vemos [porque pode ter vírus e porque pode não ser coisas próprias para a nossa idade].*
50. *Eu acho que de hoje em dia precisamos de redes sociais porque é: um bom meio de comunicação para falarmos com os nossos amigos... (hmno_2_5c_mda)*

Assinala-se, por fim, em ligação com os valores ilustrados no Gráfico 29, do registo argumentativo, que outras formas de conexão, as quais, segundo Lopes (2012: 459), têm a

⁵⁸ Relembre-se a tarefa: “Achas que as redes sociais (Facebook, Twitter, Google+, Windows Live Space, etc.) são importantes hoje em dia? Escreve um texto para ser publicado no blogue da tua escola em que exponhas a tua opinião sobre as redes sociais. Neste texto deves dizer se és a favor ou contra a existência das redes sociais. Não te esqueças de justificar a tua opinião!”

mesma função discursiva, isto é, de conetor causal de enunciação (ou justificação), como ‘visto que’, ‘dado que’, ‘uma vez que’ e ‘já que’, ou não ocorrem, ou têm uma ocorrência residual nos textos argumentativos, mesmo nos textos dos alunos do 10.º ano.

6.9. Síntese

Na tentativa de apontar para os modos como o desenvolvimento ocorre, apresentaram-se, em adição ao estudo de correlação, breves estudos de caso relativos ao uso quer do léxico, quer da sintaxe, por vezes em associação, que pudessem ser tomados como pistas mais específicas da complexidade textual. Da observação destes casos, concluiu-se, primeiramente, que a seleção lexical presente na tarefa escolar tem um forte impacto na seleção lexical realizada pelos alunos em todos os anos estudados, quer seja pela ativação de vocábulos semanticamente relacionados, quer seja pela repetição explícita, o que se comprovou com as listas de palavras mais frequentes dos textos de ambos os registos, apontando para uma clara estratégia da parte dos alunos de manutenção dos temas abordados. Uma diferença parece ser motivada pelas experiências de mundo que os alunos vivenciam, o que se notou pelo aparecimento de palavras que espelham mudanças etárias e, consequentemente, comportamentais.

A partir dos textos de registo narrativo, concluiu-se ainda que a expressão do tempo é um potencial indicador de mudanças, seja pela recategorização de funções, como acontece com o emprego do verbo ‘ir’, que, no ano inicial, tem forte uso como verbo pleno, sendo gradativamente reposicionado, até ao ano final, à condição de verbo auxiliar (aspetual temporal ou aspetual), seja pelo adensamento vocabular nas estruturas de tempo adjunto e de tempo argumental, de que se pode citar a variação lexical no escopo de adjuntos temporais construídos com o item “dia”. Ainda quanto à expressão do tempo, detetou-se que é o verbo “passar” o item lexical mais selecionado, seja flexionado, seja no particípio. É interessante pontuar que o particípio do verbo “passar” sem marcas de concordância com o sujeito gramatical é bastante recorrente em todos os anos dos textos narrativos.

Nos textos argumentativos, como demonstrado, mostrou-se relevante no processo de desenvolvimento a expressão da modalidade epistémica. Quanto à seleção lexical que a realiza, a escolha centra-se na forma verbal “achar”, apesar de, ao longo da progressão escolar, o seu emprego diminuir, dando espaço para outros verbos de valor epistémico. Os

padrões de estruturação da expressão da modalidade epistêmica com o verbo “achar” apontam também para uma variação entre os anos escolares, já que o uso da construção [achar + que] diminui em detrimento da forma [achar X + predicativo].

Por fim, demonstrou-se que a incidência de pronomes de valor anafórico se diferencia por registo, sendo predominante, no registo narrativo, o emprego de pronomes pessoais tónicos e, no argumentativo, o de demonstrativos e pessoais átonos. Em ambos os registos, notou-se que os pronomes pessoais tónicos são gradualmente abandonados no percurso do ano inicial ao ano final e, ao contrário, os átonos passam a ser mais recorrentes. Quanto à frequência de pronomes possessivos e demonstrativos, comprovou-se que há um crescimento linear ao longo da progressão escolar nos dois registos.

O uso de orações subordinadas também foi tema de estudos de caso. Quanto às orações relativas, cujo emprego se expande do 5.º ao 10.º ano, primeiramente descreveu-se a diversidade de seleção dos pronomes que as encabeçam nos textos de registo narrativo e argumentativo. Das possibilidades de pronomes, notou-se, nos dois registos e nos três anos escolares, que é o morfema *que* a escolha principal, apesar de que esta escolha se mostra decrescente do 5.º ao 10.º ano. Também nos dois registos, o pronome *que* com a função de sujeito da oração relativa é o mais acionado em todos os anos. Reparou-se, porém, que, no registo narrativo, esta função também diminui ao longo da progressão, passando o pronome *que* com função de objeto (direto) a ter ocorrências mais expressivas, ano a ano. No registo argumentativo, com uma configuração oposta, o pronome relativo *que* com função de sujeito cresce continuamente do 5.º ao 10.º ano de escolaridade, ao mesmo tempo em que as ocorrências deste pronome com função de objeto (direto) são cada vez mais reduzidas. Uma das hipóteses levantadas para este emprego diferenciado nos textos narrativos e nos textos argumentativos envolve as funções discursivas que a oração relativa desempenha no texto e, consequentemente, para o tipo de registo textual em que se enquadra. Por fim, ainda quanto ao emprego dos pronomes relativos, identificou-se que os desvios no emprego do morfema *que*, quer pela eliminação de preposição obrigatória a reger o pronome, quer pelo duplo preenchimento, em que coocorrem o pronome relativo e outro item com mesma função sintática, são mais comuns nos textos narrativos do 5.º ano, passando, a partir do 7.º ano, a terem impacto igual em ambos os registos.

Alguns usos das orações subordinadas adverbiais, pela correlação negativa que apresentaram com a progressão escolar, também foram sujeitos a descrições mais pormenorizadas. O primeiro caso de uso consistiu na descrição comparativa dos dados relativos à posição da oração adverbial (finita) em relação à sua subordinante e os dados relativos à localização temporal expressa na oração adverbial em relação à que é expressa na sua subordinante. Como resultado, percebeu-se que existe uma tendência em codificar-se a oração temporal com valor de anterioridade (em relação à situação descrita pela oração subordinante) numa posição inicial (relação anterior-inicial), embora esta seja uma tendência cada vez menos expressiva ao longo da progressão escolar. O outro caso de uso relativamente às adverbiais incidiu na diversidade de usos dos conectores que encabeçam orações subordinadas causais (em sentido lato), tendo-se constatado que a seleção dos conectores ‘porque’ e ‘pois’ corresponde, nos textos argumentativos e narrativos, à quase totalidade de ocorrências deste tipo de conectores, com a particularidade de a dinâmica destes conectores ser em alternância, isto é, enquanto no ano inicial, o conector “porque” é o mais selecionado, no ano final passa o conector “pois” a ocupar esta posição.

Os estudos de casos apresentados sobre a subordinação, corroborando os estudos de correlação, apontam para o facto de que a complexidade sintática na escrita de crianças e adolescentes em idade escolar, embora varie com o registo textual, como já notado noutras línguas por Crowhurst (1980), Verhoeven *et al.* (2002), Silva *et al.* (2010), se marca, sobretudo, nos grupos estudados, pela tendência de seleção específica de formas de realização da subordinação oracional em ambos os registos, como se comprovou, por exemplo, com a ocorrência crescente de orações relativas ou, em oposição, à ocorrência sempre decrescente de orações adverbiais.

7. Considerações finais

Nesta dissertação, apresentou-se um estudo de correlação estatística entre um conjunto de medidas de complexidade textual (lexical e sintática), amplamente referenciadas na literatura sobre o desenvolvimento linguístico, e a progressão na escolaridade básica portuguesa. Para tal, reviram-se criticamente as definições de complexidade, texto, contexto e registo, enquanto definições basilares à compreensão do objeto de investigação. Estas definições também nortearam o processo de recolha, transcrição, segmentação e anotação do *corpus*, constituído de 244 textos narrativos e argumentativos escritos por 122 alunos monolingues de português europeu inscritos no 5.º, 7.º e 10.º de escolaridade. Complementarmente ao estudo de correlação, apresentaram-se breves estudos de caso com o intuito de revelar com mais pormenor frequências expressivas de padrões de realização linguística, tanto ao nível do léxico, como da sintaxe.

No capítulo que agora se inicia, apresentam-se as considerações finais, em que primeiramente se discutem algumas das implicações pedagógicas deste trabalho, com a tentativa de enfatizar questões associadas à adoção de *corpora* como potenciais instrumentos pedagógicos e à consciência de registo, do conhecimento explícito da língua e da literacia linguística. Após isto, discutem-se algumas das limitações deste trabalho. À luz das implicações e limitações, referem-se, por fim, alguns dos desdobramentos desejados desta investigação.

7.1. Implicações pedagógicas

A primeira, e talvez a mais óbvia, implicação pedagógica do estudo que aqui se apresenta tem a ver com uma das mais potentes ideias sobre o desenvolvimento linguístico de crianças e adolescentes em idade escolar, ideia que serviu como uma das principais motivações deste estudo, que é a de que o desenvolvimento não consiste num acumular indiferenciado de recursos linguísticos, mas sim num processo caracterizado pela reorganização funcional desses recursos (Tolchinsky 2004: 234). O exame desses recursos linguísticos, nesta dissertação, fundamentou-se nos indicadores de complexidade sintática e lexical, já que revelam quantitativamente as escolhas – e, conseqüentemente, os padrões linguísticos – que os escritores realizam a partir das dimensões da língua em associação com os contextos de uso. As escolhas situadas, em que há o controlo dos recursos linguísticos de modos

diferentes dependentemente dos tipos de registos em que se inserem, refletem o que Reppen (2007: 156) chama de consciência de registo.⁵⁹

Embora os resultados dos testes de correlação aplicados aos indicadores de complexidade tenham revelado cenários de desenvolvimento estatisticamente semelhantes nos registos narrativos e argumentativos comparativamente, o exame da força da correlação, com que se demonstrou, por exemplo, que a densidade nominal é bem mais expressiva nos textos argumentativos do que nos textos narrativos, e os estudos de caso, em que se viu, por exemplo, que o emprego do pronome relativo *que* se apresenta com dinâmicas de desenvolvimento bastante divergentes num registo e noutro, permitem concluir que há da parte dos alunos investigados alguma consciência de registo (e, de novo, de padrão linguístico), ou alguma consciência de que, para narrar e argumentar, enquanto variedades linguísticas escolares, são mobilizadas configurações de uso da língua distintas para cada variedade.

Esta conclusão não é de todo surpreendente, dado que, como lembra Reppen (2007: 156), a consciência de registo se torna evidente já a partir do 3.º ano de escolaridade, mas, ainda de acordo com Reppen, é preciso considerar que esta tomada de consciência não se trata de um fenómeno estanque, já que continua a desenvolver-se nos anos de escolaridade seguintes, razão por que deve ser contínua e sistematicamente estimulada por meio de franca exposição a uma diversidade alargada de configurações linguísticas próprias de determinados registos. Neste sentido, as orientações legais para o ensino de português, em particular as que se vêem no Programa de Português (Portugal 2009), contemplam bem esta questão ao referir a necessidade de se constituir para cada ciclo escolar um “*corpus* textual”, definido como um “conjunto alargado de objectos textuais que hão-de estar presentes na aula

⁵⁹ A consciência do registo pode ser entendida como uma consequência da consciência de género, que, de modo sucinto, implica compreender criticamente os propósitos retóricos e os efeitos ideológicos das formas de género (Devitt 2004: 192).

de Português, em diversos suportes, destinados ao desenvolvimento das competências específicas quer no modo oral, quer no modo escrito” (Portugal 2009: 100).⁶⁰

Estas palavras do Programa de Português remetem a uma outra implicação pedagógica desta investigação, que está intimamente relacionada com a anterior e diz respeito ao conceito de padrão linguístico. Num sentido lato, os padrões linguísticos, como evidenciado nos estudos de correlação e de casos, apontam para uma gama vasta de constrangimentos comportamentais institucionalizados em contextos escolares (e particularmente em sala de aula), que operam sobre a produção textual dos alunos. Os alunos têm objetivos, dos quais destaco a necessidade de ser bem-sucedido. Deste modo, não é leviano afirmar que as suas produções textuais são sempre moduladas pela compreensão que possuem dos seus leitores (tipicamente o professor, que também é corretor), a quem sujeitam os seus textos (tal como a si próprios). Desta forma, os padrões linguísticos são para os alunos estratégias para atingir tais objetivos. A forte manutenção vocabular ancorada no vocabulário das tarefas de escrita ou a seleção de uso limitada a dois conetores causais, cenários recorrentes em todos os anos escolares, exemplificam tais estratégias. Como é sabido, os professores de português, pela sua contínua experiência em sala de aula, têm intuições a respeito destas estratégias. No entanto, estudos baseados em *corpora*, como aqui se fez, permitem a observação e a documentação metódica das ocorrências que conformam os padrões linguísticos, permitindo superar um desafio que se coloca a muitos professores, que é o de ter a segurança sistemática de saber como textos de diferentes registos são particularizados linguisticamente, isto é, quais padrões de uso os diferenciam entre si.

Portanto, uma implicação pedagógica evidente deste estudo consiste na defesa do recurso a *corpora* enquanto fonte regular para o planeamento de programas, matrizes e aulas e ainda para a elaboração de gramáticas e manuais didáticos. *Corpora* de textos escritos por adultos escolarizados mostram-se como instrumentos cruciais para a observação e análise dos usos reais que se fazem da língua, servindo, assim, como uma referência válida para a legitimação ou não de formas de realização possíveis na variante culta. Ter a consciência,

⁶⁰ De notar que no Programa e Metas Curriculares de Português do Ensino Básico, homologado em 2015, não há mais qualquer referência ao conceito de *corpus* textual, como se via adequadamente no programa homologado em 2009.

com o devido rigor científico, de que a expressão do tempo com o verbo “passar” no particípio sem marcas de concordâncias (como em “passado dois anos”, “passado uma semana”, etc., recorrente nos textos escolares estudados) é um fenómeno frequente na escrita adulta escolarizada pode, de um lado, assegurar aos professores de português a possibilidade de tratar esse fenómeno não mais como desviante. De outro lado, pode também dar-lhes a hipótese justificada de combater a sua existência na língua. De uma forma ou de outra, troca-se a mera intuição por dados estatísticos.

Quanto aos *corpora* desenvolvimentais (longitudinais ou *quasi*-longitudinais), do mesmo modo que os *corpora* de língua adulta escolarizada, também eles agregam valor científico ao conhecimento empírico que professores (e todos os outros agentes educacionais) têm do percurso do desenvolvimento linguístico de crianças e adolescentes em idade escolar. São também ricos instrumentos para a elaboração de programas e metas curriculares, podendo contribuir para que estes sejam abalizados pelas particularidades reais do desenvolvimento, o que permitiria compreender, por exemplo, a insistente semelhança nos números estatísticos identificados nos estudos de correlação (lexical e sintático) entre os grupos de alunos do 5.º e do 7.º ano, que, apesar da aproximação etária, são pertencentes a ciclos diferentes da educação básica.

Não se ignora naturalmente que, ao contrário do que acontece para a língua inglesa, em português ainda são poucos os *corpora* disponíveis *online*. Quanto a *corpora* desenvolvimentais, após pesquisa, não se identificou qualquer um disponível *online*, falha que se pretende colmatar com a disponibilização futura do *corpus* utilizado nesta investigação. Quanto a *corpora* de português adulto e escolarizado, são bons exemplos o Corpus de Referência do Portuguese Contemporâneo⁶¹, o CetemPúblico⁶², CorpusEye⁶³ e o Corpus de Português⁶⁴. Embora poucos, são eficientes o bastante para garantir a vantagem de se recorrer

⁶¹ <http://alfclul.clul.ul.pt/CQPweb/crpcf15/>

⁶² <http://acdc.linguateca.pt/cetempublico/whatisCETEMP.html>

⁶³ <http://corp.hum.sdu.dk/cqp.pt.html>

⁶⁴ www.corpusdoportugues.org/

a eles a fim de desconstruir a noção muito comum, pelo menos entre alunos, de que a gramática é um conjunto imutável de regras e independente dos contextos de produção, contribuindo decisivamente para salientar o caráter tanto determinístico, como probabilístico dos padrões linguísticos.

O domínio de padrões linguísticos, que se constituem por vocábulos e estruturas prováveis ou determinadas, parece ser uma questão francamente afetada pelo desenvolvimento da literacia, em particular da literacia linguística. Entende-se por literacia linguística, de acordo com Ravid e Tolchinsky (2002: 422), a posse de um repertório de estruturas linguísticas ricas e variadas, que permite realizar uma vasta gama de registos e géneros e que se adapta a diversos objetivos e contextos comunicativos. Nesta visão de literacia linguística, o papel da escrita assume-se como central, o que fica evidente nas palavras de Olson (2006, cf. Chik 2015: 340): “a set of social practices that exploit the affordances of writing for particular ends”. Desta afirmação, também se reforça que as propriedades potenciais que a escrita pode propiciar a um escritor – *affordances*⁶⁵, como exemplifica Olson (2006, cf. Berman 2009: 102): um vocabulário mais expandido, uma gramática mais formalizada ou variadas estruturas retóricas – a conduzem a ser usadas com certos tipos de efeitos, relativamente previsíveis e, acrescento, conscientemente motivados, em certos tipos de contextos, como se vê nos seus usos académicos ou literários. Deste modo, para a instanciação eficiente das propriedades da escrita, torna-se necessário associá-la ao conhecimento explícito da língua, a qual se substancia através de instruções formais e sistémicas – por vezes amparadas em metalinguagem – sobre as suas unidades e o seu funcionamento.

⁶⁵ Conforme Paiva (2012: 2), o termo “affordance”, cunhado por James Gibson (1986), teve sua origem nos estudos sobre ecologia, “para nomear alguma coisa que se referisse tanto ao ambiente quanto ao animal, implicando a complementaridade entre o animal e o ambiente.” Norman (1990: 9) propõe outra definição (mais próxima da aceção utilizada por Olson 2006): “the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used”. Embora seja hoje uma expressão recorrente em várias áreas do saber, como *design* industrial ou inteligência artificial, parece não ter encontrado ainda uma tradução assente no português, sendo frequentemente utilizada no original, conforme se vê, por exemplo, no dicionário multilingue *online* Linguee (www.linguee.pt, com acesso em 01/08/2015).

O exame de frequências e rácios de realização léxico-gramatical em textos narrativos e argumentativos obtidos em contexto natural de sala de aula, de que resultaram textos completos e genuínos, permite afiliar esta investigação à perspectiva da escrita como produto⁶⁶, o que é coerente com a definição de texto enquanto atualização linguística situada (tópico discutido no Capítulo 2), pelo que abre espaço para que se façam algumas reflexões sobre a relação entre o ensino de escrita e o ensino de gramática, relação que tem tido lugar cativo, como foco de discussão, no cenário da linguística aplicada à educação há pelo menos duas décadas, quer em Portugal (Costa 2006, 2010; Duarte 1991, 1997, 2004; Martins e Niza 2001; Teixeira 2014; Teixeira e Santos 2011; Teixeira *et al.* 2015), quer no estrangeiro (Perera 1984; Martin e Rothery 1993; Cassany 1996, 1999; Hudson 2001; Andrews 2006; Jones *et al.* 2013; Myhill 2012, 2013, 2014). Desta discussão, resulta a convicção de que o conhecimento explícito da língua funciona como um “facilitador da mobilização de recursos linguísticos (...) em situações de escrita”, como bem sintetiza Costa (2010: 289). Nesta tese, portanto, a par dos autores citados, levanta-se mais uma voz a favor do ensino de gramática enquanto parte fundamental da construção da literacia linguística dos alunos da educação básica.

Já constitui um posicionamento claro em documentos diretores nacionais que a gramática e a escrita são eixos complementares no processo global do ensino, como se vê, por exemplo, no Programa de Português para o Ensino Básico (Portugal 2009: 15-6), em que se atesta que o conhecimento explícito da língua, “dependente do ensino explícito, formal e sistematizado e sendo transversal” às competências de escrita, “permite o controlo das regras e a selecção dos procedimentos mais adequados à compreensão e à expressão, em cada situação comunicativa.”⁶⁷. No entanto, alguns desafios continuam presentes. Como lembra

⁶⁶ Esta perspectiva complementa-se pela perspectiva da escrita como processo, isto é, como atividade cognitiva complexa, baseada em estratégias de solução de problemas e processos de composição (Johnson e Johnson 1999: 342).

⁶⁷ No novo programa de português do Ensino Básico (Portugal 2015), mantém-se a mesma linha de raciocínio sobre a relação entre a gramática e a escrita, constituindo-se como um dos objetivos a serem atingidos ao fim da escolaridade básica a mobilização dos “conhecimentos gramaticais para aperfeiçoar as capacidades de interpretar e produzir enunciados orais e escritos” (2015: 6) No entanto, em alguns pareceres publicitados no período da consulta pública ao novo programa, são contundentes em

Costa (2010: 7), estão em exercício professores de português de diferentes gerações e, consequentemente, de diferentes formações, o que resulta em visões de ensino (tanto da gramática, como da escrita) também diferentes, não parecendo suficiente a “mera integração de tais conteúdos no processo formativo” desses professores (Pereira *et al.* 2014: 131). É preciso, portanto, refletir sobre os processos de transposição didática de conteúdos gramaticais às práticas de escrita diversificadas para as salas de aula. É, deste modo, que a adoção de *corpora* de referência e desenvolvimentais, bem como um maior estímulo à consciência registral tornam-se grandes aliados nestes processos de didatização e de literacia.

Apesar de ser quase incontestável que uma criança que cresce num ambiente monolíngue tem acesso à maioria das estruturas morfológicas e sintáticas da sua língua antes de entrar na escola, procurou-se reforçar, neste estudo, a ideia de que há evidências de que o desenvolvimento da língua é um processo prolongado, que não se encerra aos dez ou doze anos, e de que consideráveis mudanças em todos os domínios linguísticos também acontecem na língua de crianças mais velhas e adolescentes, e não apenas na primeira infância (cf. Nippold, 1992; Berman e Verhoeven, 2002; Berman e Katzemberger 2004; Ravid e Tolchinsky, 2002). Isto, de facto, torna a língua dos adultos diferente da língua dos adolescentes, e a dos adolescentes diferente da língua de uma criança de doze anos. É, portanto, ao longo desse período da infância tardia e da adolescência que grande parte das características

criticar a redução dos conteúdos a serem ensinados como parte do eixo da competência gramatical. No parecer do Centro de Linguística da Universidade do Porto critica-se a “ausência de qualquer referência a: adjetivo relacional, relações temporais como a anterioridade, posterioridade e simultaneidade de situações”. No parecer da Associação dos Professores de Português (APP), está referida a incoerência entre os objetivos indicados no novo programa e as metas curriculares do domínio da Gramática, do 1.º ao 9.º ano, que exigem sobretudo “tarefas classificatórias aos alunos, com descritores de desempenho e objetivos que implicam que os professores realizem inúmeros exercícios de identificação e de reconhecimento de categorias, de integração de elementos em classes, de aplicação de regras, num claro apelo à memorização, sem reinvestimento cognitivo no desenvolvimento da leitura, da escrita e da oralidade”. O primeiro parecer referido está disponível em: <http://www.app.pt/6560/parecer-do-centro-de-linguistica-da-universidade-do-porto-sobre-a-proposta-de-novo-programa-de-portugues-do-ensino-basico/> e o segundo, em: <http://www.app.pt/6547/17-de-abril-de-2015-parecer-sobre-metas-e-programa-do-ensino-basico/>.

da língua adulta emerge e se consolida, acompanhada de construções de natureza mais complexa, cujas funções textuais são próprias de determinados tipos de textos a que só se tem acesso por meio da educação formal.

7.2. Limitações

Os resultados e as conclusões deste estudo devem ser entendidos à luz de um conjunto de limitações, as quais, em seus potenciais desdobramentos, podem servir de baliza para reconsiderações. Grande parte delas relaciona-se com o procedimento metodológico de recolha e anotação do *corpus*. A extensão original do *corpus* (575 textos, produzidos por 287 alunos), após a aplicação dos critérios de exclusão, resultou numa quantidade muito abaixo do inicial (244 textos, 122 alunos). Apesar de se ter a informação sobre quais os alunos que não eram falantes monolíngues de português europeu antes da aplicação dos estímulos, não foi possível fazer a triagem atempadamente. Um dos pontos cruciais deste estudo centrava-se na obtenção de textos produzidos por alunos de um determinado ciclo que refletissem a aprendizagem do ciclo imediatamente anterior, razão por que todo o processo de obtenção ocorreu ainda nos três primeiros meses do ano letivo, a fim de se impedir influências significativas do novo ciclo de estudo. Acrescente-se a isto o facto de terem sido poucas as escolas participantes do estudo. Vale a pena lembrar que, mesmo com a aquiescência do Ministério da Educação, do total de treze *e-mails* enviados aos diretores de agrupamentos e escolas, apenas sete responderam, dos quais somente quatro concordaram em participar.

Outra possível limitação deste estudo, ainda relacionada com o *corpus*, foi a ausência de um *subcorpus* de controlo. Os resultados encontrados nos textos escritos pelos alunos dos três anos de escolaridade, se comparados com um *subcorpus* de escrita de adultos já escolarizados que tivessem recebido tratamento computacional semelhante, tornaria este trabalho mais informativo, em particular sobre os resultados identificados nos textos dos alunos do 10.º ano.

Também, o processo de segmentação e anotação de orações e unidades-t, apesar de todo o esforço aplicado no sentido de uniformizá-lo, é passível de falhas, já que ocorreu de modo completamente manual. Em muitos momentos, houve incertezas sobre a classificação a adotar. No entanto, é de salientar que muitas destas incertezas ecoam controvérsias que

se veem na literatura linguística de um modo geral, e, por vezes, representam questões bastante complexas, como é o tratamento usualmente dicotómico entre coordenação e subordinação.

7.3. Desdobramentos desta investigação

É expectável que este estudo tenha desdobramentos, nomeadamente em relação ao *corpus*, que tem potencialidades para ser disponibilizado *online* (obviamente com o resguardo absoluto do anonimato das escolas e dos alunos participantes, já assumido na realização deste estudo) transformando-se numa fonte sistemática de consultas, tanto para novas investigações sobre o desenvolvimento linguístico, como para a preparação de aulas, materiais didáticos e matrizes curriculares regulados por padrões de uso. A disponibilização *online* depende de um conjunto de ações que reforcem a fiabilidade das informações linguísticas, o que inclui a revisão por um grupo alargado de investigadores tanto da anotação automática ao nível do léxico, como da segmentação e anotação manual ao nível da sintaxe. É enriquecedor o acréscimo de uma camada de anotação que considere a representação de elementos da estrutura textual, preferencialmente em linha com diretrizes internacionais (*Text Encoding Initiative Guidelines*, por exemplo) de modo a propiciar a comparação com outros *corpora*, inclusive de línguas diferentes. Também, para a disponibilização *online*, é importante a ampliação do *corpus*, com a inclusão de novas escolas participantes e de novos registos escolares (descritivo ou injuntivo, por exemplo), transformando-o, deste modo, num *corpus* monitor (Sinclair 1991, cf. McEnery e Hardie 2012), que garantirá, portanto, uma maior representatividade da população estudada.

Em paralelo com a disponibilização do *corpus*, pretende-se realizar, como continuidade desta investigação, trabalhos de aprofundamento, o que se fará pela adoção de outras técnicas de medição de complexidade lexical, como, por exemplo, a verificação da densidade de itens lexicais por especialidade (nomes, verbos, adjetivos e advérbios terminados em *-mente*) mensurado com base em diferentes unidades linguísticas (orações ou unidades-t) e também da densidade de nomes abstratos, que, segundo Berman e Ravid (2008: 100), consiste numa faceta da diversidade lexical que só emerge tardiamente na escrita, e sempre em função da idade e da escolaridade. Na dimensão sintática, buscar-se-á corroborar, por meio da aplicação de testes estatísticos, se os achados descritos nos estudos de caso, como,

por exemplo, a diversificação no emprego de conectores oracionais causais, estão matematicamente correlacionados com a progressão escolar.

Também é uma das intenções de continuidade desta investigação realizar um estudo de correlação em que se institua como variável independente a qualidade da escrita, que, baseada na avaliação dos textos do *corpus* por um conjunto de professores do ensino básico, permitirá identificar que características linguísticas podem ser tomadas como indicadoras de uma escrita considerada por estes profissionais como de melhor qualidade. Abordagem semelhante, muito recorrente em particular no cenário académico anglófono, vê-se, por exemplo, nos estudos de Crowhurst (1980), Jordan (1993), Allison *et al.* (2002) ou McNamara *et al.* (2010).

Este estudo aponta ainda para a necessidade de que, em estudos futuros, refinamentos destes indicadores, como a diversidade baseada em lemas, ou outros indicadores de desenvolvimento da complexidade lexical, como a raridade ou a extensão das palavras, por exemplo, sejam aplicados a *corpora* de desenho semelhante e maiores, a fim de se atestar, por exemplo, se a ausência de diferenças entre o 5.º e o 7.º ano se mantém ou não e, não se mantendo, qual é a natureza qualitativa da mudança. Por fim, embora limitado ao exame da complexidade em textos produzidos por crianças e adolescentes nos estádios típicos de escolarização, este estudo também pode fornecer informações relevantes para a pesquisa sobre o desenvolvimento atípico desse grupo populacional, podendo inclusive ser tomado como um referencial para os mecanismos diagnósticos de dificuldades linguísticas ou mesmo de dificuldades de aprendizagem.

Os desdobramentos referidos acima, do mesmo modo que este estudo, representam apenas uma pequena parte das possibilidades de exame do desenvolvimento linguístico em idade escolar, cuja dimensão escrita, ainda pouco explorada, estabelecem por si só uma agenda de investigação prolongada. Naturalmente, esta agenda enriquecer-se-á com novas pesquisas levadas a cabo por outros investigadores, baseadas em novos *corpora*, mais extensos, mais diversificados quanto aos registos e, conseqüentemente, mais representativos, que possam refutar ou validar as conclusões desta investigação, construindo coletivamente um corpo de conhecimento sobre os modos como as crianças e os adolescentes nas aulas de português usam a língua e como isto (se) reflete (nas) práticas e pedagogias de ensino da escrita em português.

.

8. Referências bibliográficas

- Abadiano, H.R. (1995) Cohesion strategies and genre in expository prose: an analysis of the writing of children of ethnolinguistic cultural groups. *Pragmatics*, 5, 299-324.
- Adam, J.M. (1995) Hacia una definición de la secuencia argumentativa. *Comunicación, Lengua y Educación*, 25, 9-22.
- Allison, P., Beard, R. & Willcocks, J. (2002) Subordination in children's writing. *Language in Education*, 16, 97-111.
- Almeida, M.J.S.D. (2013) Orações adverbiais temporais: desenvolvimento linguístico e construção de texto narrativo. Mestrado, Universidade Nova de Lisboa.
- Altenberg, B. (1984) Causal linking in spoken and written english. *Studia Linguistica*, 38, 20-69.
- Andrews, R. (2006) The effect of grammar teaching on writing development. *British Educational Research Journal*, 32, 39-55.
- Anglin, J.M. (1993) Vocabulary development: a morphological analysis. Wiley, Nova Iorque.
- Antunes, J.D. & Brito, A.M. (2008) Contribuição para a definição do perfil linguístico dos alunos do ensino básico: o caso das orações relativas. In *Actas do Colóquio de Homenagem a Fernanda Irene Fonseca*, (Eds, Oliveira, F. & Duarte, I.M.) pp. 237-254.
- Asher, N. (1993) Reference to abstract objects in discourse. Kluwer Academic Publishers, Boston.
- Auer, P. (1996) From context to contextualization. *Links & Letters*, 3, 11-28.
- Avent, J.R. & Austermann, S. (2003) Reciprocal scaffolding: a context for communication treatment in aphasia. *Aphasiology*, 17, 397-404.
- Baker, P. & Ellege, S. (2011) Key terms in discourse analysis. Continuum, Londres/Nova Iorque.
- Baker, P., Hard, A. & McEnery, T. (2006) A glossary of corpus linguistics. Edinburgh University Press, Edimburgo.
- Bamberg, B. (1984) Assessing coherence: A reanalysis of essays written for the National Assessment of Educational Progress, 1969-1979. *Research in the Teaching of English*, 18, 305-319.
- Barbosa, P. (2013) Subordinação argumental finita. In *Gramática do Português*, (Eds, Raposo, E.B.P. et al.) Fundação Calouste Gulbenkian, Lisboa, pp. 1821-1900.
- Barbosa, P. & Raposo, E.B.P. (2013) Subordinação argumental infinitiva. In *Gramática do Português*, (Eds, Raposo, E.B.P. et al.) Fundação Calouste Gulbenkian, Lisboa, pp. 1901-1980.
- Barry, S. & Lazarte, A. (1998) Evidence for mental models: how do prior knowledge, syntactic complexity, and reading topic affect inference generation in a recall task for nonnative readers of Spanish? *Modern Language Journal*, 82, 176-193.
- Bawarshi, A.S. & Reiff, M.J. (2010) Genre: an Introduction to history, theory, research, and pedagogy. Parlor Press/The WAC Clearinghouse, West Lafayette.
- Beaman, K. (1984) Coordination and subordination revisited: syntactic complexity in spoken and written narrative discourse. In *Coherence in spoken and written discourse*, (Ed, Tannen, D.) Ablex, Norwood, pp. 45-80.
- Beaugrande, R. (2011) Text linguistics. In *Discursive pragmatics*, (Eds, Zienkowski, J., Östman, J.O. & Verschueren, J.) John Benjamins, Amsterdão/Filadélfia, pp. 286-296.
- Beaugrande, R. & Dressler, W. (1981) Introduction to text linguistics. Longman, Londres.
- Bechara, E. (2009) Moderna gramática portuguesa. Nova Fronteira, Rio de Janeiro.
- Beers, S.F. & Nagy, W.E. (2009) Syntactic complexity as a predictor of adolescent writing quality: which measures? which genre? *Reading and writing*, 24, 185-200.
- Beers, S.F. & Nagy, W.E. (2011) Writing development in four genres from grades three to seven: syntactic complexity and genre differentiation. *Reading and writing*, 24, 183-202.
- Berber-Sardinha, T. (2004) Linguística de corpus. Manole, São Paulo.
- Berber-Sardinha, T. & Shimazumi, M. (2003) Schoolchildren writing: a corpus-based analysis. *Linguagem e Ensino*, 6, 11-33.

- Bereiter, C. & Scardamalia, M. (1982) From conversation to composition: the role of instruction in a developmental process. Lawrence Erlbaum Associates, Hillsdale NJ.
- Berman, R.A. (2000) Children's innovative verbs vs nouns: structured elicitations and spontaneous coinages. In *Methods for studying language production*, (Eds, Menn, L. & Bernstein-Ratner, N.) Erlbaum, Mahwah, NJ, pp. 69-93.
- Berman, R.A. (2004) The role of context in developing narrative abilities. In *Relating events in narrative: typological and contextual perspectives: volume 2*, (Eds, Stromqvist, S. & Verhoeven, L.) Laurence Erlbaum, Nova Jersey/Londres, pp. 261-280.
- Berman, R.A. (2007) Developing linguistic knowledge and language use across adolescence. In *Blackwell handbook of language development*, (Eds, Hoff, E. & Shatz, M.) Blackwell, Malden/Oxford/Victoria, pp. 347-367.
- Berman, R.A. & Katzemberger, I. (2004) Form and function in introducing narrative and expository texts. *Discourse Processes*, 38, 57-94.
- Berman, R.A. & Nir, B. (2010) The lexicon in writing-speech-differentiation: developmental perspectives. *Written Language and Literacy*, 13, 183-205.
- Berman, R.A. & Ravid, D. (2009) Becoming a literate language user: oral and written text construction across adolescence. In *Cambridge handbook of literacy*, (Eds, Olson, D.O. & Torrance, N.) Cambridge University Press, Cambridge, pp. 92-111.
- Berman, R.A. & Verhoeven, L. (2002) Cross-linguistic perspectives on the development of text-production abilities: speech and writing. *Written Language and Literacy*, 5, 1-44.
- Berninger, V.W., Nagy, W. & Beers, S. (2011) Child writers' construction and reconstruction of single sentences and construction of multi-sentence texts: contributions of syntax and transcription to translation. *Reading and writing*, 24, 151-182.
- Bernstein, B. (1960) Language and social class. *The British Journal of Sociology*, 11, 271-276.
- Bernstein, B. (1971) Class, codes and control: theoretical studies toward a sociology of language. Routledge, Londres.
- Bernstein, B. (1972) Social class, language and socialization. In *Language and Social Context*, (Ed, Giglioli, P.P.) Penguin, Harmondsworth, pp. 157-178.
- Bernstein, B. (1999) Vertical and horizontal discourse: an essay. *British Journal of Sociology of Education*, 20, 157-173.
- Biber, D. (1992) On the complexity of discourse complexity: a multidimensional analysis. *Discourse Processes*, 15, 133-163.
- Biber, D. (1995) Dimensions of register variation: a cross linguistic comparison. Cambridge University Press, Cambridge.
- Biber, D. (2006) University language: a corpus-based study of spoken and written registers. John Benjamins, Amesterdão/Filadélfia.
- Biber, D. & Conrad, S. (2009) Register, genre, and style. Cambridge University Press, Cambridge.
- Biber, D., Gray, B. & Poonpon, K. (2011) Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45, 5-35.
- Biber, D. et al. (2007) Longman grammar of spoken and written english. Longman, Londres.
- Bick, E. (2000) The parsing system Palavras - automatic grammatical analysis of portuguese in a constraint grammar framework. Doutorado, Aarhus University.
- Bloor, T. & Bloor, M. (2004) The functional analysis of english. Arnold, Londres.
- Blount, N.S., Fredrick, W.C. & Johnson, S.L. (1968) The effect of a study of grammar on the writing of eighth-grade students. Wisconsin Research and Development Center for Cognitive Learning, Madison, Wisconsin.
- Bos, P. (2001) Temporality issues in moroccan arabic and dutch. In *Narrative development in a multilingual context*, (Eds, Verhoeven, L. & Stromqvist, S.) John Benjamins, Amesterdão/Filadélfia.
- Brainerd, B. (1982) The type-token relation in the works of S. Kierkegaard. In *Computing in the humanities*, (Ed, Bailey, R.W.) North-Holland, Amesterdão, pp. 97-109.
- Brito, A.M. (2003) Subordinação adverbial. In *Gramática da Língua Portuguesa*, (Eds, Mateus, M.H.M. et al.) pp. 695-728.
- Brown, G. & Yule, G. (1983) Discourse analysis. Cambridge University Press, Cambridge.

- Bucks, R.S. et al. (2000) Analysis of spontaneous, conversational speech in dementia of Alzheimer type: evaluation of an objective technique for analyzing lexical performance. *Aphasiology*, 14, 71-91.
- Bügel, K. & Buunk, B.P. (1996) Sex differences in foreign language text comprehension: the role of interests and prior knowledge. *Modern Language Journal*, 80, 15-30.
- Bussmann, H. (2006) *Routledge dictionary of language and linguistics*. Routledge, Londres/Nova Iorque.
- Butler, C. (2003) *Structure and function: a guide to three major structural-functional theories. Volume 2: From clause to discourse and beyond*. Benjamins, Amsterdão.
- Carlisle, J.F. (2000) Awareness of the structure and meaning of morphologically complex words: impact on reading. *Reading and Writing*, 12, 169-190.
- Carroll, J.B. (1964) *Language and thought*. Englewood Cliffs, Nova Jersey.
- Carvalho, J.A.B.S. (1990) A evolução sintática na produção escrita de crianças e adolescentes. *Revista Portuguesa de Educação*, 3, 29-13.
- Carvalho, J.A.B.S. (2003) *Escrita: percursos de Investigação*. DME/UM, Minho.
- Cassany, D. (1996) *Describir el escribir. Cómo se aprende a escribir*. Editorial Paidós, Barcelona.
- Cassany, D. (1999) *Construir la escritura*. Editorial Paidós, Barcelona.
- Cavaco, A.M. & Várzea, D. (2010) Contribuição para o estudo da leitura de folhetos informativos nas farmácias portuguesas. *Rev Port Saúde Pública*, 28, 179-186.
- Cavique, L. (2008) Legibilidade de artigos científicos: análise de dados da RCC. *Revista de Ciências da Computação*, 3,
- Chafe, W. (1984) How people use adverbial clauses. *Berkeley Linguistics Society*, 10, 437-449.
- Cheung, H. & Kemper, S. (1992) Competing complexity metrics and adults' production of complex sentences. *Applied Psycholinguistics*, 13, 53-76.
- Chik, A. (2015) Popular culture, digital words and second language learners. In *The Routledge handbook of literacy studies*, (Eds, Rowsell, J. & Pahl, K.) Routledge, Londres, pp. 339-353.
- Chipere, N. et al. (2001) Using a corpus of school children's writing to investigate the development of vocabulary diversity. *Proceedings of the Corpus Linguistics Conference*, 126-133.
- Choupina, C.M.G.M. (2010) Orações relativas: como e quando inseri-las na sala de aula? *Exedra: Revista Científica*, 1, 49-68.
- Christiansen, T. (2011) *Cohesion: a discourse perspective*. Peter Lang, Berna.
- Christie, F. (1986) Writing in schools: generic structures as ways of meaning. In *Functional approaches to writing: research perspectives*, (Ed, Couture, B.) Frances Pinter, Londres, pp. 221-240.
- Cloran, C. (1999) Context, material situation and text. In *Text and context in functional linguistics*, (Ed, Ghadessy, M.) John Benjamins, Amsterdão/Filadélfia,
- Cock, S. (2010) Spoken learner corpora and efl teaching. In *Corpus-Based Approaches to English Language Teaching*, (Ed, Campoy-Cubillo, M.C., Bellés-Fortuño, B., and Gea-Valor, M. L) Continuum, Londres/Nova Iorque, pp. 123-137.
- Colleta, J.M. (2004) *Le développement de la parole chez l'enfant âgé de 6 à 11 ans: corps, langage et cognition*. Pierre Mardaga, Sprimont.
- Colombi, M.C. (2002) Academic language development in latino student's writing. In *Developing advanced literacy in first and second languages*, (Eds, Schleppegrell, M.J. & Colombi, M.C.) Lawrence Erlbaum Associates, Mahwah, pp. 67-86.
- Colwell, K., Hiscock, C.K. & Memon, A. (2002) Interviewing techniques and the assessment of statement credibility. *Applied Cognitive Psychology*, 16, 287-300.
- Cornish, F. (2009) Text and discourse as context: discourse anaphora and the fdg contextual component. *Web Papers in Functional Discourse Grammar*, 82, 97-115.
- Costa, A.L. (2006) *Desenvolvimento linguístico e literacia de escrita*. Actas do Encontro Internacional de Linguística Aplicada, Digital.
- Costa, A.L. (2010) *Estruturas contrastivas: desenvolvimento do conhecimento explícito e da competência de escrita*. Doutorado, Universidade de Lisboa.
- Costa, J. et al. (2009) *Produção e compreensão de orações relativas em português europeu: da*

dos do desenvolvimento típico, de PEDL e do agramatismo. XXIV Encontro Nacional da APL, 211-224.

Crespo-Allende, N. & Alfaro-Faccio, P. (2010) Desarrollo tardío del lenguaje: la conciencia metapragmática en la edad escolar. *Universitas Psychologica*, 9, 229-240.

Crossley, S.A., Varner, L.K. & McNamara, D.S. (2013) Cohesion-based prompt effects in argumentative writing. *Proceedings of the 26th International Florida Artificial Intelligence Research Society Conference*, 202-207.

Crossley, S.A. et al. (2011) The development of writing proficiency as a function of grade level: a linguistic analysis. *Written Communication*, 28, 282-311.

Crowhurst, M. (1979) The effect of syntactic on teacher's quality ratings of narration and arguments.

Crowhurst, M. (1980) Syntactic complexity in narration and argument at three grade levels. *Canadian Journal of Education*, 5, 6-13.

Crowhurst, M. (1987) Cohesion in argument and narration at 3 grade levels. *Research in the Teaching of English*, 21, 185-201.

Crowhurst, M. & Piche, G.L. (1979) Audience and mode of discourse effects on syntactic complexity in writing at two grade levels. *Research in the Teaching of English*, 13, 101-109.

Crystal, D. (2008) *Dictionary of linguistics and phonetics*. Blackwell, Malden/Oxford/Victoria.

Cunha, C. & Cintra, L. (2010) *Nova gramática do português contemporâneo*. Lexikon, Rio de Janeiro.

Cunha, L.F. (2000) Alguns valores temporais das orações com quando. XV Encontro Nacional da Associação Portuguesa de Linguística, 311-333.

Daelemans, W. et al. (1996) MBT: A memory-based part of speech tagger generator. In *Fourth workshop on very large corpora*, (Eds, Ejhed, E. & Dagan, I.) pp. 14-27.

Dahl, O. (2004) The growth and maintenance of linguistic complexity. *John Benjamins, Amsterdão/Filadélfia*.

Dahl, O. (2007) Definitions of complexity. *Proceedings of the Colloquium on complexity, accuracy and fluency in second language use, learning & teaching*, 41-46.

Dahl, O. (2009) Testing the assumption of complexity invariance: the case of elfdalian and Swedish. In *Language complexity as an evolving variable*, (Eds, Sampson, G., Gil, D. & Trudgill, P.) Oxford University Press, Oxford, pp. 50-63.

Daller, H., Milton, J. & Treffers-Daller, J. (2007) *Modelling and assessing vocabulary knowledge*. Cambridge University Press, Cambridge.

Dancey, C. & Reidy, J. (2004) *Statistics without Maths for Psychology: using SPSS for Windows*. Prentice Hall, Londres.

Devitt, A. (2004) *Writing genres*. Carbondale, Southern Illinois UP.

Diessel, H. (2001) The ordering distribution of main and adverbial clauses: a typological study. *Language*, 77, 343-365.

Diessel, H. (2005) Competing motivations for the ordering of main and adverbial clauses. *Linguistics*, 43, 449-470.

Diessel, H. (2008) Iconicity of sequence: a corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics*, 19, 465-490.

Dittmer, A.E. (1971) A comparison of the grammatical structures of professional senior high and junior high expository writing. *Doutoramento, DAI*.

Dixon, E. (1970) Syntactic indexes and student writing performance. *Doutoramento, University of Chicago*.

Dixon, R.M.W. (2006) Complement clauses and complementation strategies in typological perspective. In *Complementation: A cross-linguistic typology*, (Eds, Dixon, R.M.W. & Aikhenvald, A.Y.) Oxford University Press, Oxford, pp. 1-48.

Duarte, I. (1991) Ensinar português: para quê e como? *Palavras*, 11, 66-74.

Duarte, I. (1991) Funcionamento da língua: a periferia dos NPP. In *Documentos do Encontro sobre os Novos Programas de Português*, (Ed, Delgado-Martins, M.R.) Colibri, Lisboa, pp. 45-59.

Duarte, I. (2003) Aspectos linguísticos da organização textual. In *Gramática da Língua Portuguesa*, (Eds, Mateus, M.H.M. et al.) pp. 85-123.

Duarte, I. & Brito, A.M. (2003) Predicação e classe de predicados verbais. In *Gramática da Língua Portuguesa*, (Eds, Mateus, M.H.M. et al.) pp. 179-204.

- Duarte, I. & Costa, A.L. (2004) The place of portuguese as first language in the National Curriculum of Basic Education: essential competences. In *Flexibility in Curriculum. Citizenship and Communication*, (Ed, Zabalza, M.) Ministério da Educação, Lisboa, pp. 323-330.
- Durán, P. et al. (2004) Developmental trends in lexical diversity. *Applied Linguistics*, 25, 220-242.
- Edmonds, B. (1999) Syntactic measures of complexity. Doutoramento, University of Manchester.
- Eggins, S. (1994) *An introduction to systemic functional linguistics*. Pinter, Londres.
- EUA. (2010) Common core state standards for english language arts and literacy in history/social studies, science, and technical subjects: research supporting key elements of the standards. National Governors Association for Best Practices & Council of Chief State School Officers,
- Everitt, B.S. & Skrondal, A. (2010) *The cambridge dictionary of statistics*. Cambridge University Press, Cambridge.
- Fetzer, A. (2004) *Recontextualizing context*. John Benjamins, Amesterdão/Filadélfia.
- Figueiredo, O.M. (2000) *A anáfora nominal em textos de alunos: a língua no discurso*. Doutoramento, Universidade do Porto.
- Fisher, D., Frey, N. & Lapp, D. (2012) *Text complexity: raising rigor in reading*. International Reading Association, Newark.
- Fitzgerald, J. & Spiegel, D. (1986) Textual cohesion and coherence in children's writing. *Research in the Teaching of English*, 20, 263-280.
- Fletcher, P. & Garman, M. (1986) *Language acquisition*. Cambridge University Press, Cambridge.
- Fodor, J.A., Bever, T. & Garrett, M. (1974) *The psychology of language*. McGraw-Hill, Nova Iorque.
- Foley, W. & van Valin, R. (1984) *Functional syntax and universal grammar*. Cambridge University Press, Cambridge.
- Ford, C.E. (1993) *Grammar in interaction: adverbial clauses in American English conversations*. Cambridge University Press, Cambridge.
- Foster, P. & Skehan, P. (1996) The influence of planning and task type on second language performance. *Studies in second language acquisition*, 18, 299-323.
- Gálan Rodríguez, C. (1999) La subordinación causal y final. In *Gramática Descriptiva de la Lengua Española*, (Eds, Bosque, I. & Demonte, V.) Espasa-Calpe, Madrid, pp. 3597-3643.
- Généreux, M., Hendrickx, I. & Mendes, A. (2012) A large portuguese corpus cleaning and pre-processing. *Computational Processing of the Portuguese Language. Proceedings of the 10th International Conference PROPOR1012*, 113-120.
- Généreux, M., Hendrickx, I. & Mendes, A. (2015) *Manual for the CRPC CQPweb interface. Manual 1.0 Version February, 2011*, 9.
- Gibson, J.J. (1986) *The ecological approach to visual perception*. Lawrence Erlbaum Associates, Hillsdale.
- Gillis, S. & Ravid, D. (2009) Language acquisition. In *Cognition and pragmatics*, (Eds, Dominiek, S., Östman, J. & Verschueren, J.) John Benjamins, Amesterdão/Filadélfia, pp. 201-249.
- Givón, T. (1991) Markedness in grammar: distributional, communicative and cognitive correlates of syntactic structure. *Studies in Language*, 15/2, 335-370.
- Gonçalves, A. & Costa, T. (2002) (Auxiliar a) *Compreender os verbos auxiliares*. APP, Colibri, Lisboa.
- Gonçalves, F.M.R., Guerreiro, P. & Freitas, M.J. (2011) *O conhecimento da língua: percursos de desenvolvimento*. Ministério da Educação. Direção-Geral de Inovação e de Desenvolvimento Curricular, Lisboa.
- Granger, S. (2008) Learner corpora. In *Corpus linguistics an international handbook*, (Eds, Lüdeling, A. & Kytö, M.) Walter de Gruyter, Berli/Nova Iorque,
- Greenbaum, S. (1996) *The oxford english grammar*. Oxford University Press, Oxford.
- Guiraud, P. (1960) *Problèmes et méthodes de la statistique linguistique*. D. Reidel, Dordrecht.
- Gundel, J.K., Hedber, N. & Zacharski, R. Cognitive status and the form of referring expressions in discourse. *Language*, 69, 175-204.
- Halliday, M.A.K. (1978) *Language as social semiotic*. Edward Arnold, Londres.

- Halliday, M.A.K. (1987) Spoken and written modes of meaning. In *Comprehending oral and written language*, (Eds, Horowitz, R. & Samuels, S.J.) Academic Press, Orlando,
- Halliday, M.A.K. (1989) *Spoken and written language*. Oxford University Press, Oxford.
- Halliday, M.A.K. (2005) The spoken language corpus: a foundation for grammatical theory. In *Computational and Quantitative Studies*, (Ed, Webster, J.J.) Continuum, Londres/Nova Iorque, pp. 157-190.
- Halliday, M.A.K. (2007) The notion of “context” in language education. In *Language and education*, (Eds, Halliday, M.A.K. & Webster, J.J.) Continuum, Londres/Nova Iorque,
- Halliday, M.A.K. (2009) Methods - techniques - problems. In *Continuum companion to systemic functional linguistics*, (Eds, Halliday, M.A.K. & Webster, J.J.) Continuum, pp. 59-86.
- Halliday, M.A.K. & Hasan, R. (1989) *Language, text and context: aspects of language in a social-semiotic perspective*. Oxford University Press, Oxford.
- Halliday, M.A.K. & Matthiessen, C. (2004) *Introduction to functional grammar*. Edward Arnold, Londres.
- Halliday, M.A.K., McIntosh, A. & Stevens, P. (1964) *The linguistic sciences and language teaching*. Longman, Londres.
- Halliday, M.A.K. & Webster, J.J. (2009) *Continuum companion to systemic functional linguistics*. Continuum,
- Harpin, W.S. (1973) *Social and educational influences on children’s acquisition of grammar*. Social Science Research Council, Swindon.
- Harpin, W.S. (1976) *The second ‘R’ : writing development in the junior school*. Allen and Unwin, Londres.
- Hasan, R. (1995) The conception of context in text. In *Discourse in society: systemic functional perspectives - meaning and choice in language: studies for Michael Halliday*, (Eds, Fries, P. & Gregory, M.) Ablex, Norwood, pp. 183 - 283.
- Hasan, R. (1996) What’s going on: a dynamic view of context in language. In *Ways of saying: ways of meaning*, (Eds, Butt, D., Cloran, C. & Williams, G.) Cassel, Londres, pp. 37-50.
- Hasan, R. (1999) Speaking with reference to context. In *Text and context in functional linguistics*, (Ed, Ghadessy, M.) Benjamins, Amesterdão, pp. 219-328.
- Hasan, R. (2009) The place of context in a systemic functional model. In *Continuum companion to systemic functional*, (Eds, Halliday, M.A.K. & Webster, J.J.) Continuum, Londres/Nova Iorque,
- Hewings, A. & Hewings, M. (2005) *Context and grammar: an advanced resource book*. Londres/Nova Iorque: Routledge. Routledge, Londres/Nova Iorque.
- Hinkel, E. (2006) Current perspectives on teaching the four skills. *TESOL Quarterly*, 40, 109-131.
- Huddleston, R. & Pullum, G. (2002) *The cambridge grammar of the english language*. Cambridge University Press, Cambridge.
- Huddleston, R. & Pullum, G. (2006) Coordination and subordination. In *The Handbook of english linguistics*, (Eds, Aarts, B. & McMahon, A.) Blackwell Publishing, Malden/Oxford/Victoria, pp. 198-219.
- Hudson, R. (2008) Linguistic theory. In *The handbook of educational linguistics*, (Eds, Spolski, B. & Hult, F.M.) Blackwell, Malden MA/Oxford, pp. 53-65.
- B, H.S. et al. (2003) *Delineando a pesquisa clínica: uma abordagem epidemiológica*. Artmed, Porto Alegre.
- Hunt, K.W. (1965) *Grammatical structures written at three grade levels*. National Council of Teachers of English, Champaign.
- Hunt, K.W. (1970) *Syntactic maturity in schoolchildren and adults*. Chicago University Press, Chicago.
- Hyland, K. (2004) *Second language writing*. Cambridge University Press, Nova Iorque.
- Hyland, K. & Paltridge, B. (2011) *Continuum companion to discourse analysis*. Continuum, Londres/Nova Iorque.
- Jisa, H. (2004) Growing into academic french. In *Language development across childhood and adolescence*, (Ed, Berman, R.A.) John Benjamins, Amesterdão/Filadélfia, pp. 135-162.
- Jisa, H. & Kern, S. (1998) Relative clauses in french children’s narratives. *Journal of Child Language*, 25, 623-652.
- Johansson, V. (2009) *Developmental aspects of text production in writing and speech*. Doutorado, Lund University.

- Johnson, K. & Johnson, H. (1999) *Encyclopedic dictionary of applied linguistics: a handbook for language teaching*. Blackwell, Oxford/Malden.
- Jones, R.H. & Lock, G. (2011) *Functional grammar in the esl classroom: noticing, exploring and practicing*. Palgrave Macmillan, Londres.
- Jones, S. et al. (2013) Playful explicitness with grammar: a pedagogy for writing. *Literacy*, 47, 103-111.
- Jordan, M. (1993) Towards an understanding of mature writing: analyzing and paraphrasing complex noun phrases. *Technostyle*, 11, 39-72.
- Karmiloff-Smith, A. (1986) Some fundamental aspects of language development after age 5. In *Language acquisition*, (Eds, Fletcher, P. & Garman, M.) Cambridge University Press, Cambridge, pp. 455-474.
- Kupersmitt, J. (2004) Form-function relations in spanish narratives: comparative study of bilingual and monolingual speakers. In *Relating events in narrative*, (Eds, Strömquist, S. & Verhove, L.) Lawrence Erlbaum, Mahwah/Londres, pp. 395-434.
- Kusters, W. (2003) *Linguistic complexity: the influence of social change on verbal inflection*. LOT, Utrecht.
- La Brant, L. (1933) A study of certain language development of children in grades 4-12 inclusive. *Genetic Psychology Monographs*, 14, 387-491.
- Labov, W. (1972) *Sociolinguistic patterns*. University of Pennsylvania Press, Filadélfia.
- Larsen-Freeman, D. & Cameron, L. (2008) *Complex systems and applied linguistics*. Oxford University Press, Oxford.
- Laws, G. & Gunn, D. (2002) Relationships between reading, phonological skills and language development in individuals with Down syndrome: a five year follow up study. *Reading and Writing*, 15, 527-548.
- Lee, D. (2001) Genres, registers, text types, domains, and styles: clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology*, 5, 37-72.
- Leech, G. (1992) Corpora and theories of linguistic performance. *Directions in corpus linguistics* 105-122. In *Directions in corpus linguistics: proceedings of the Nobel Symposium*, (Ed, Svartvik, J.) Mouton de Gruyter, Berlim.
- Linell, P. (1998) *Approaching dialogue talk, interaction and contexts in dialogical perspectives*. John Benjamins., Amesterdão/Filadélfia.
- Linell, P. (2005) *The written language bias in linguistics: its nature, origins, and transformation*. Routledge, Londres.
- Litosseliti, L. (2010) *Research methods in linguistics*. Continuum, Londres/Nova Iorque.
- Loban, W.D. (1963) *The language of elementary school children: a study of the use and control of language and the relations among speaking, reading, writing, and listening*. NCTE, Champaign.
- Loban, W.D. (1976) *Language development: kindergarten through grade twelve*. NCTE, Urbana.
- Lobo, M. (2001) Para uma sintaxe das orações causais do português. *Actas do XVI Encontro Nacional da APL*, 291-306.
- Lobo, M. (2003) Aspectos da sintaxe das orações subordinadas adverbiais. *Doutorado, Universidade Nova de Lisboa*.
- Lobo, M. (2013) Subordinação adverbial. In *Gramática do Português*, (Eds, Raposo, E.B.P. et al.) Fundação Calouste Gulbenkian, Lisboa, pp. 1981-2060.
- Longacre, R.E. (1983) *The grammar of discourse*. Plenum, Londres/Nova Iorque.
- Lopes, A.C.M. (2012) Contributos para uma análise semântico-pragmática das causais de enunciação no português europeu contemporâneo. *Alfa*, 56, 451-468.
- Lopes, H.C. (2004) Aspectos sintáticos, semânticos e pragmáticos das construções causais: contributo para uma reflexão sobre o ensino da gramática. *Doutoramento, Universidade do Porto*.
- López Garcia, Á. (1999) Relaciones paratáticas e hipotáticas. In *Gramática Descriptiva de la Lengua Española*, (Eds, Bosque, I. & Demonte, V.) Espasa-Calpe, Madrid, pp. 3507-3549.
- MacLure, M. (1999) Home and school language. In *Concise encyclopedia of educational linguistics*, (Ed, Spolski, B.) Pergamon/Elsevier, Oxford, pp. 202-203.
- MacWhinney, B. (2000) *The Childes Project: tools for analyzing talk*. Lawrence Erlbaum Associates, Mahwah.
- Malvern, D. et al. (2004) *Lexical diversity and language development: quantification and assessment*. Palgrave Macmillan, Nova Iorque.

- Marques, R. (2013) Construções de grau. In Gramática do Português, (Eds, Raposo, E.B.P. et al.) Fundação Calouste Gulbenkian, Lisboa, pp. 2141-2173.
- Martin, J.R. (1997) Register and genre: modelling social context in functional linguistics - narrative genres. First Lisbon International Meeting on Discourse Analysis, 305-344.
- Martin, J.R. & Rose, D. (2003) Working with discourse: meaning beyond the clause. Continuum, Londres.
- Martin, J.R. & Rothery, J. (1993) Grammar: making meaning in writing. In The powers of literacy: a genre approach to teaching writing, (Eds, Cope, B. & Kalantzis, M.) Falmer Press, Londres, pp. 137-154.
- Martin, J.R. (2001) Cohesion and texture. In The handbook of discourse analysis, (Eds, Schiffrin, D., Tannen, D. & Hamilton, H.E.) Blackwell, Malden/Oxford, pp. 35-53.
- Martins, M. (2008) A contribuição dos significados experienciais para a constituição dos elementos obrigatórios do género narrativa escolar: um estudo de caso. Mestrado, Universidade de Lisboa.
- Martins, M.A. & Niza, I. (2001) Psicologia da aprendizagem escrita. Universidade Aberta, Lisboa.
- Matos, G. & Brito, A.M. (2002) On the syntax of canonical comparatives in european portuguese. Journal of Portuguese Linguistics, 1, 41-81.
- Matos, G. & Raposo, E.B.P. (2013) Estruturas de coordenação. In Gramática do Português, (Eds, Raposo, E.B.P. et al.) Fundação Calouste Gulbenkian, Lisboa, pp. 1759-1816.
- Matos, G. (2003) Estruturas de coordenação. In Gramática da Língua Portuguesa, (Eds, Mateus, M.H.M. et al.) pp. 549-592.
- Mazur-Palandre, A. & Jisa, H. (2010) Le flux de l'information: une analyse développementale de la preferred argument structure (Du Bois, 1987, 2003). 2ème Congrès Mondial de Linguistique Française,
- McCarthy, P.M. & Jarvis, S. (2010) MTLTD, vocd-D, and HD-D: a validation study of sophisticated approaches to lexical diversity assessment. Behavior Research Methods, 42, 381-392.
- McCutchen, D. & Perfetti, C. (1982) Coherence and connectedness in the development of discourse production. Text, 2, 113-139.
- McEnery, T. & Hardie, A. (2012) Corpus linguistics: method, theory, and practice. Cambridge University Press, Cambridge.
- McNamara, D.S., Crossley, S.A. & McCarthy, P.M. (2010) Linguistic features of writing quality. Written Communication, 27, 57-86.
- Mendes, A. (2013) Organização textual e articulação de orações. In Gramática do Português, (Eds, Raposo, E.B.P. et al.) Fundação Calouste Gulbenkian, Lisboa, pp. 1691-1758.
- Mendes, M. (2015) Para um modelo ontogénico da escrita infantil no 1.º ciclo do Ensino Básico: a emergência do período como unidade de organização textual. Doutoramento, Universidade de Lisboa.
- Móia, T. (1992) A sintaxe das orações relativas sem antecedente expresso do português. Mestrado, Universidade de Lisboa.
- Móia, T. (1996) A sintaxe das orações relativas sem antecedente expresso do português: uma abordagem segundo a teoria dos princípios e parâmetros. In Quatro estudos em sintaxe do português, (Ed, Ferro, F.M.) Colibri, Lisboa, pp. 149-188.
- Móia, T. (2003) Subdomínios de significação associados ao tempo: uma panorâmica geral. In Razões e emoção. Miscelânea de estudos em homenagem a Maria Helena Mira Mateus, (Eds, Castro, I. & Duarte, I.) INCM, Lisboa, pp. 95-127.
- Móia, T. & Alves, A.T. (2013) Tempo ad-junto e tempo discursivo. In Gramática do Português, (Eds, Raposo, E.B.P. et al.) Fundação Calouste Gulbenkian, Lisboa, pp. 557-584.
- Morin, E. (2002) Educação e complexidade: os sete saberes e outros ensaios. Cortez, São Paulo.
- Myhill, D. & Fisher, R. (2010) Writing development: cognitive, sociocultural, linguistic perspectives. Journal of Research in Reading, 33, 1-3.
- Neubert, A., and Shreve, G. M. (1992) Translation as text. Kent State University Press, Kent.
- Nippold, M.A. (2004) Research on later language development. In Language development across childhood and adolescence, (Ed, Berman, R.A.) John Benjamins, Amesterdão/Filadélfia: pp. 1-8.
- Nippold, M.A., Hesketh, L.J. & Duthie, J.K. (2005) Conversational versus expository discourse: a study of syntactic development in children, adolescents, and adults. Journal of Speech, Language, and Hearing Research, 48, 1048-1064.

- Nippold, M.A., Mansfield, T.C. & Billow, J.L. (2007) Peer conflict explanations in children, adolescents, and adults: Examining the development of complex syntax. *American Journal of Speech-Language Pathology*, 179-188.
- Nippold, M.A., Schwarz, I.E. & Undlin, R. (1992) Use and understanding of adverbial conjuncts: a developmental study of adolescents and young adults. *Journal of Speech and Hearing Research*, 35, 108-118.
- Norman, D.A. (1990) The design of everyday things. Doubleday, Nova Iorque.
- O'Donnell, R.C., Griffin, W.J. & Norris, R.C. (1967) Syntax of kindergarten and elementary school children: a transformational analysis. NCTE, Champaign.
- O'Donnell, M. (2007) The UAM Corpus Tool. ICAME,
- Oliveira, F. (2003) Tempo e aspecto. In *Gramática da Língua Portuguesa*, (Eds, Mateus, M.H.M. et al.) pp. 127-178.
- Oliveira, F. & Mendes, A. (2013) Modalidade. In *Gramática do Português*, (Eds, Raposo, E.B.P. et al.) Fundação Calouste Gulbenkian, Lisboa, pp. 623-672.
- Olson, D. (2006) Oral discourse in a world of literacy. *Research in Teaching of English*, 41, 136-142.
- Paiva, V.L.M. (2010) O propiciamento (affordance) e autonomia na aprendizagem de língua inglesa. (Ed, Lima, D.C.) Edições UESB, Vitória da Conquista, pp. Digital.
- Pereira, I.S.P. (2008) Para a caracterização do contexto de ensino - aprendizagem da literacia no 1.º ciclo de escolaridade: das competências dos alunos às concepções e práticas dos professores. Doutoramento, Universidade do Minho.
- Pereira, L.A., Graça, L. & Carnin, A. (2014) Modelos de formação para o ensino da escrita em Portugal e no Brasil. *Educação e contemporaneidade*, 23, 129-137.
- Perera, K. (1984) Children's writing and reading: analysing classroom language. Blackwell, Londres.
- Perera, K. (1986) Grammatical differentiation between speech and writing in children aged 8-12. In *The Writing of writing*, (Ed, Wilkinson, A.) Milton Keynes, Londres, pp. 90-108.
- Peres, J.A. (1997) Sobre conexões proposicionais em português. In *Sentido que a vida faz: estudos para Oscar Lopes*, (Ed, Brito, A.M.) Campo das Letras, Porto, pp. 775-787.
- Peres, J.A. & Mascarenhas, S. (2006) Notes on sentential connections (predominantly) in Portuguese. *Journal of Portuguese Linguistics*, 5, 113-169.
- Peres, J.A. & Móia, T. (1995) Áreas críticas da língua portuguesa. Caminho, Lisboa.
- Perini, M. (2010) Gramática do português brasileiro. Parábola, São Paulo.
- Piot, M. (1988) Coordination-subordination: une définition générale. *Langue Française*, 77, 5-18.
- Polio, C. (2000) Research methodology in second language writing research: the case of text-based studies. In *On second language writing*, (Eds, Silva, T. & Matsuda, P.K.) Lawrence Erlbaum Associates, Mahwah/Londres,
- Ponte, J.P. (2006) Estudos de caso em educação matemática. *Bolema*, 25, 105-132.
- Portugal. (2009) Programa de português do ensino básico. Ministério da Educação e Ciência. Direcção-Geral da Educação, Lisboa.
- Portugal. (2012) Metas curriculares de português - ensino básico: 1.º, 2.º e 3.º ciclos. Ministério da Educação e Ciência. Direcção-Geral da Educação, Lisboa.
- Portugal. (2015) Programa e metas curriculares de português do ensino básico. Ministério da Educação e Ciência. Direcção-Geral da Educação, Lisboa.
- Pyykkönen, P. & Järviö, J. (2012) Children and situation models of multiple events. *Institute for Psycholinguistics*, 1-33.
- Quirk, R. et al. (1985) A comprehensive grammar of the English language. Longman, Londres/Nova Iorque.
- Ramsay, V. (1987) The functional distribution of preposed and postposed if and when clauses in written discourse. In *Coherence and grounding in discourse*, (Ed, Tomlin, R.) John Benjamins, Amsterdão, pp. 383-408.
- Raposo, E.B.P. (2013) Estrutura da frase. In *Gramática do português*, (Eds, Raposo, E.B.P. et al.) Fundação Calouste Gulbenkian, Lisboa, pp. 303-400.
- Ravid, D. (2004) Emergence of linguistic complexity in later language development: evidence

from expository text construction. In *Perspectives on language and language development: essays in honor of Ruth A. Berman*, (Eds, Ravid, D.D. & Shyldkrot, H.B.) Kluwer Academic Publishers, Dordrecht/Boston/Londres,

Ravid, D. & Berman, R.A. (2006) Information density in the development of spoken and written narratives in english and hebrew. *Discourse Processes*, 41, 117-149.

Ravid, D. & Tolchinsky, L. (2002) Developing linguistic literacy: a comprehensive model. *Journal of Child Language*, 29, 419-448.

Read, J. (2000) *Assessing vocabulary*. Cambridge University Press, Cambridge.

Renzi, L. & Salvi, G. (1981) *Grande grammatica italiana di consultazione*. Il Mulino, Bologna,

Reppen, R. (2007) First language and second language writing development of elementary students. In *Corpus-based perspectives in linguistics*, (Eds, Kawaguchi, Y. et al.) John Benjamins, Amesterdão/Filadélfia, pp. 147-168.

Reppen, R., Fitzmaurice, S.M. & Biber, D. (2002) *Using corpora to explore linguistic variation*. John Benjamins, Amesterdão/Filadélfia.

Richards, B.J. & Malvern, D.D. (1997) *Quantifying lexical diversity in the study of language development*. University of Reading, Reading.

Richards, J.C. & Schmidt, R. (2002) *Longman dictionary of applied linguistics and language teaching*. Longman, Londres.

Rodrigues, S.B.P. (2008) *Escrita espontânea: desenvolvimento das capacidades de composição escrita em crianças do 1º ao 4º ano de escolaridade*. Mestrado, Universidade Fernando Pessoa.

Ruth, L. & Murphy, S. (1988) *Designing writing tasks for the assessment of writing*. Greenwood, Mahwah.

Sampson, G. (2005) *The Lucy corpus documentation*. <http://www.grsampson.net/LucyDoc.html>.

Scarton, C. & Aluísio, S.M. (2010) *Análise da inteligibilidade de textos via ferramentas de processamento de língua natural: adaptando as métricas do Coh-Metrix para o português*. *LinguaMática*, 2, 45-62.

Schleifer, M. (2003) *Development of written text production of native israeli and ethiopian immigrant school children and adolescent*. Doutorado, Tel-Aviv University.

Schlepppegrell, M.J. (1992) Subordination and linguistic complexity. *Discourse Processes*, 15, 117-131.

Schlepppegrell, M.J. (2002) Challenges of the science register for esl students: errors and meaning-making. In *Developing advanced literacy in first and second languages*, (Eds, Schlepppegrell, M.J. & Colombi, M.C.) Lawrence Erlbaum, Mahwah/Nova Jérsei/Londres, pp. 119-142.

Schlepppegrell, M.J. (2004) *The language of schooling: a functional linguistics perspective*. Lawrence Erlbaum, Mahwah/Londres.

Scott, C.M. (2004) Syntactic contributions to literacy learning. In *Handbook of language and literacy: development and disorders*, (Eds, Stone, C.A. et al.) The Guilford Press, Nova Iorque, pp. 340-362.

Shanahan, T. (2012) *Text complexity*. McGraw-Hill Reading Wonders: Research Base Alignment, 1-4.

Sichel, H.S. (1986) Word frequency distributions and type-token characteristics. *Mathematical Scientist*, 11, 45-72.

Silva, M.L., Sánchez Abchi, V. & Borzone, A. (2010) Subordinated clauses usage and assessment of syntactic maturity: a comparison of oral and written retellings in beginning writers. *Journal of Writing Research*, 2, 47-64.

Sinclair, J. (2004) *Trust the text: language, corpus and discourse*. Routledge, Londres/Nova Iorque.

Smith, J.A. & Kelly, C. (2002) Stylistic constancy and change across literary corpora: using measures of lexical richness to date works. *Computers and the Humanities*, 36, 411-430.

Stoddard, S. (1990) *Text and texture: patterns of cohesion*. Ablex, Mahwah.

Strömqvist, S. et al. (2001) Toward a cross-linguistic comparison of lexical quanta in speech and in writing. *Written Language and Literacy*, 5, 46-67.

Strömqvist, S., Nordqvist, A. & Wengelin, A. (2004) Writing the frog story: developmental and cross-modal perspectives. In *Relating events in narrative*, (Eds, Strömqvist, S. & Verhoven, L.) Lawrence Erlbaum, Mahwah/Londres, pp. 359-394.

Stubbs, M. (1996) *Text and corpus analysis*. Blackwell, Oxford.

Sun, L. & Nippold, M.A. (2012) *Narrative writing in children and adolescents: examining the literate lexicon*. American Speech-Language-Hearing Association, 43,

Sweetser, E. (1999) From etymology to pragmatics. Cambridge University Press, Cambridge.

Teixeira, M. (2011) Metas de Aprendizagem e Novos Programas de Português: uma leitura do pré-Escolar ao 3.º CEB. In Novos desafios no ensino do português, (Eds, Teixeira, M., Silva, I. & Santos, L.) Escola Superior de Educação de Santarém, Santarém, pp. 11-19.

Teixeira, M. (2014) Relación entre gramática y escritura en el currículo de lengua Portuguesa. *Bel-laterra Journal of Teaching & Learning Language & Literature*, 7, 1-20.

Teixeira, M., Gorgulho, A.R. & Lopes, S. (2014) Gramática e escrita - que relação? I Congresso Internacional sobre la Enseñanza de la Gramática - presente, passat i futur, Digital.

Templin, M. (1957) Certain language skills in children: their development and inter-relationships. University of Minnesota Press, Minnesota.

Thompson, S.A. (1985) Grammar and written discourse: initial and final purpose clauses in english. *Quantified Studies in Discourse*, 5, 55-84.

Tolchinsky, L. (2004) The nature and scope of later language development. In *Language development across childhood and adolescence*, (Ed, Berman, R.A.) John Benjamins, Amesterdão/Filadélfia, pp. 233-248.

Tutin, A. (2002) A corpus-based study of pronominal anaphoric expressions in french. *Discourse Anaphora and Anaphora Resolution*, Proceedings.

Ure, J. (1971) Lexical density and register differentiation. In *Applications of linguistics*, (Eds, Perren, G.E. & Trim, I.L.M.) Cambridge University Press, Cambridge, pp. 443-452.

Valente, P. (2008) Produção de frases relativas em alunos do terceiro ciclo do ensino básico e do ensino secundário. Mestrado, Universidade de Lisboa.

van den Bosch, A. & Daelemans, W. (1999) Memory-based morphological analysis. 37th Annual Meeting of the Association for Computational Linguistics, 285-292.

van Dijk, T. (1976) Text and context. Longman, Nova Iorque.

van Dijk, T. (2007) Macro contexts. In *Discourse and international relations*, (Eds, Scheu, D. & Sanchéz, J.S.) Peter Lang, Berna, pp. 3-26.

van Dijk, T. (2008) Discourse and context. Cambridge University Press, Cambridge.

van Valin, R. (2004) An introduction to syntax. Cambridge University Press, Cambridge.

Vasconcelos, M. (1991) Compreensão e produção de frases com orações relativas. Um estudo experimental com crianças dos três anos e meio aos oito anos. Mestrado, Universidade de Lisboa.

Vasconcelos, M. (1996) Compreensão e produção de frases relativas em português europeu. In *Introdução à linguística geral e portuguesa*, (Ed, Faria, I.H.U.) Caminho, Lisboa, pp. 323-330.

Veloso, R. (2013) Subordinação relativa. In *Gramática do Português*, (Eds, Raposo, E.B.P. et al.) Fundação Calouste Gulbenkian, Lisboa, pp. 2061-2136.

Verhoeven, L. et al. (2002) Clause packaging in writing and speech: a cross-linguistic developmental analysis. *Written Language and Literacy*, 5, 135-161.

Verhoeven, L. & Stromqvist, S. (2001) Development of narrative production in a multilingual context. In *Narrative development in a multilingual context*, (Eds, Verhoven, L. & Stromqvist, S.) John Benjamins, Amesterdão/Filadélfia, pp. 1-14.

Vila, I. (1991) El desarrollo tardío del lenguaje. *La enseñanza de la lengua*, 7-12.

Wagner, R.K. et al. (2011) Modeling the development of written language. *Reading and writing*, 24, 203-220.

Webster, J.J. (2009) The essential Halliday. Continuum, Londres/Nova Iorque.

Widdowson, H.G. (1973) An applied linguistic approach to discourse analysis. Doutoramento, University of Edinburgh.

Widdowson, H.G. (1978) Teaching language as communication. Oxford University Press, Oxford.

Widdowson, H.G. (2004) Text, context, pre-text. Blackwell, Malden/Oxford/Victoria.

Wilkinson, A. et al. (1980) Assessing language development. Oxford University Press, Oxford.

Wodak, R. (2011) Continuum companion to discourse analysis. (Eds, Hyland, K. & Paltridge, B.) Continuum, Londres/Nova Iorque.

Woese, C.R. (2004) A new biology for a new century. *Microbiology and Molecular Biology Reviews*, 68, 173-186.

Wolfe-Quintero, K., Inagaki, S. & Kim, H.Y. (1998) Second language development in writing:

measures of fluency, accuracy, and complexity. University of Hawai'i at Mānoa, Honolulu.

Woods, N. (2000) New zealand english across generations: an analysis of selected vowel and consonant variables. In New zealand english, (Eds, Bell, A. & Kuiper, K.) Victoria University Press, Wellington, pp. 84-110.

Wray, A. et al. (2001) Projects in linguistics: a practical guide to researching language.

Wray, D. & Medwell, J. (2006) Progression in writing and the Northern Ireland

levels for writing. University of Warwick, Warwick.

Yde, P. & Spoelders, M. (1985) Text Cohesion - an exploratory-study with beginning writers. Applied Psycholinguistics, 6, 407-415.

9. ANEXOS

9.1. Protocolo de cooperação

A Direção do Agrupamento de Escolas _____ e o investigador Mário Martins celebram um Protocolo de Cooperação para o desenvolvimento de um trabalho de investigação científica sobre a caracterização do perfil linguístico de crianças e jovens que tenham concluído o 1.º, 2.º ou 3.º ciclos do Ensino Básico em Portugal.

Ao abrigo deste protocolo, o Agrupamento de Escolas _____ aceita acolher o projeto de investigação proposto por aquele investigador para a realização de uma tese de doutoramento a apresentar à Faculdade de Letras da Universidade de Lisboa.

O trabalho de investigação no Agrupamento de Escolas _____ deve ocorrer no primeiro semestre do ano letivo 2011/12, sendo o marco inicial a assinatura deste Protocolo de Cooperação. Nesse período, o investigador irá recolher produções escritas de _____ turma(s) do 5.º ano e/ou _____ turma(s) do 7.º ano e/ou _____ turma(s) do 10.º ano produzidas em sala de aula a partir de estímulos definidos pelo investigador. Esses textos constituirão um *corpus* que será objeto de uma análise científica.

O investigador compromete-se a fornecer ao Agrupamento de Escolas _____ uma cópia, em formato digital, da tese de doutoramento resultante desta investigação, depois de aprovada pelo júri, e também cópias, em formato digital, de todos os artigos científicos publicados resultantes da referida investigação.

O investigador compromete-se a manter o Agrupamento de Escolas _____, seus professores e alunos autores dos textos em absoluto anonimato, bem como a respeitar outras obrigações de sigilo que venham a ser acordadas entre o Agrupamento de Escolas _____ e o investigador.

Lisboa, ____ de _____ de 2011

Mário Martins (investigador)

Agrupamento de Escolas _____

9.2. Termo de consentimento

Termo de consentimento informado e de permissão para a utilização de registos de produções textuais do/a educando/a:

Conforme a circular informativa que lhe foi enviada em ____/____/ 2011, a Direção do Agrupamento de Escolas _____ aceitou acolher a investigação conduzida pelo investigador Mário Martins, no âmbito de um projecto de doutoramento em Linguística, na Faculdade de Letras da Universidade de Lisboa, sob orientação dos Professores Doutores Rui Marques e Telmo Mória, com o apoio da Fundação para a Ciência e Tecnologia (FCT).

A investigação tem como objetivo a caracterização linguística, do ponto de vista longitudinal, de produções textuais de crianças e jovens que tenham concluído um dos ciclos da escolaridade básica em Portugal. Essa investigação terá lugar no primeiro semestre do ano lectivo de 2011/12.

O desenvolvimento da investigação implica a aplicação de estímulo apropriado e a recolha para registo, em suporte digital, de produções escritas dos alunos produzidas, em contexto regular de sala de aula, em datas especificadas em acordo entre o investigador e a Direção.

Os registos escritos serão utilizados para fins de constituição de um *corpus* de produções textuais e a sua divulgação em publicações de carácter científico salvaguardará o absoluto anonimato dos respetivos autores.

Autorização do/a Encarregado/a de Educação:

Eu, _____, autorizo a utilização dos registos das produções textuais do meu/minha educando/a no âmbito do projeto científico acima descrito.

Assinatura do(a) Encarregado(a) de Educação: _____

Local e Data: _____

9.3. Questionário para a identificação da L1 dos participantes

BLOCO 1: Informações Gerais

Iniciais do Nome: Turma:

Idade:

Local de Nascimento:

Sexo:

BLOCO 2: Informações Específicas

Lá em casa, em que língua falas com o teu(s) encarregado(s) de educação? (Podes marcar mais de uma opção)

Português

☐

Português brasileiro

☐

Português africano

☐

Outras línguas

☐

Qual foi a tua média a Português no ano anterior?

_____ valores.

9.4. Variáveis de investigação

DIMENSÃO LINGÜÍSTICA	INDICADOR	VARIÁVEIS	CÓDIGO
Aplicáveis a todas as dimensões linguísticas	Aplicáveis a todos os indicadores	ano escolar	ano
Complexidade lexical	Diversidade	rácio do número de itens diferentes pelo número total de itens	typtok
	Densidade	rácio do número de itens lexicais pelo número total de itens	lextok
Complexidade sintática	Extensão	frequência média de palavras por oração	tokorc
		frequência média de palavras por unidade-t	tokunt
		frequência média de orações por unidade-t	orcunt
	Variação	rácio do número de orações subordinadas pelo número total de orações subordinadas	suborc
		rácio do número de orações subordinadas completivas pelo número total de orações subordinadas	comsub
		rácio do número de orações subordinadas adverbiais pelo número total de orações subordinadas	advsb
		rácio do número de orações subordinadas relativa pelo número total de orações subordinadas	relsub
		rácio do número de orações subordinadas gradativas pelo número total de orações subordinadas	grasub

9.5. Valores absolutos

	5.º ano	7.º ano	10.º ano	Total
Valores relativos à complexidade lexical				
<i>tokens</i>	8.673	15.354	19.658	43.685
<i>types</i>	4.945	9.021	12.138	26.104
<i>tokens</i> lexicais	3.878	6.922	8.893	19.693
<i>tokens</i> gramaticais	4.795	8.432	10.765	23.992
Valores relativos à complexidade sintática				
<i>tokens</i>	8.673	15.354	19.658	43.685
unidades-t	888	1.481	1.567	3.936
unidades-t simples	525	860	791	2.176
unidades-t multioracionais	363	621	776	1.760
orações (sub + sim)	1.517	2.578	3.034	7.129
orações subordinadas	629	1.097	1.467	3.193
orações completivas	248	410	583	1.241
orações adverbiais	254	433	433	1.120
orações relativas	124	251	435	810
orações gradativas	3	3	16	22

9.6. Orientações gerais de segmentação e anotação de orações

As orientações que a seguir se apresentam, de caráter geral, nem sempre são controversas e podem, por vezes, contrariar orientações já consideradas tradicionais. Estas regras, que não se pretendem exaustivas, baseiam-se num conjunto de decisões que tenta, na medida do possível, conciliar as proposições da linguística moderna (e das gramáticas descritivas) com as especificidades de uma língua em desenvolvimento, como é a língua escrita escolar aqui investigada.

1. Se um conjunto de orações ligadas entre si por coordenação, suplementação ou justaposição funcionar como argumento de um predicator, salvo exceções (como se verá adiante), somente o primeiro membro – o mais contíguo – se encaixa no segmento que contém o predicator e, conseqüentemente, somente ele é anotado como uma oração encaixada, com a classificação a depender do tipo de seleção. As orações seguintes são segmentadas à parte, tendo o estatuto de segmento independente do primeiro segmento. Esta proposta de segmentação sustenta-se no facto de as orações seguintes poderem ser realizadas autonomamente:

51. [Chegamos à casa da minha melhor amiga,] [disse-lhe: [- Vamos embora]] [estamos atrasadas] [ela disse: [- Só são 12h da tarde]] [vamos chegar lá às 15h30m.] [- Ok vamos] [estou ansiosa [para começar as férias.]] [- Eu também] [disse ela.] (dsra_1_5h_mda)

Acima, têm-se dois compostos por justaposição. O primeiro constitui-se pelas orações [Vamos embora] e [estamos atrasadas]. Em conformidade com a regra descrita, apenas o primeiro membro da combinação recebe a anotação, neste caso de completiva. O mesmo acontece com a combinação [Só são 12h da tarde] [vamos chegar lá às 15h30m.].

Constituem exceções à regra acima:

- casos em que o conector de subordinação está explícito na(s) oração(ões) seguinte(s):

52. [Há acontecimentos [que partilhamos com certas pessoas] [e que marcam a nossa vida para sempre]]. (cacafm_1_10_fdl)

- casos em que o núcleo da oração seguinte é um verbo não-finito com elipse do verbo auxiliar (mas expresso na primeira oração):

53. [O Facebook é uma rede social [em que podemos falar com os amigos,] [partilhar coisas que gostamos]]

- casos em que o núcleo verbal da oração seguinte está no conjuntivo e o composto oracional funciona como sujeito:

54. [[Por exemplo quem tem Facebook] [e passe muito tempo a surfar nele] pode ter de usar óculos.]

2. Apesar de as orações relativas apositivas, tradicionalmente conhecidas como adjetivas explicativas, constituírem *per se*, como afirma Móia (2013: 2067), "um grupo sintático e prosódico autónomo", a opção aqui é tratá-las à semelhança das relativas, isto é, incluí-las no segmento onde está o sintagma nominal a que estão ligadas. As orações relativas de comentário, que tipicamente retomam estruturas de natureza frásica e que são iniciadas pelo conjunto 'o que', recebem o mesmo tratamento.

3. As orações adverbiais, pelo seu carácter de não-argumento, são segmentadas separadamente das orações de que são dependentes, como ilustrado abaixo:

55. [Quando vi a notícia] [fiquei em pânico,] [não sabia [em que autocarro é que ela ia,] [nem como estava.]] (dsv_1_10a_hnn)

4. Outra regra diz respeito à elipse. Para a descrição da regra aqui adotada, veja-se primeiramente o exemplo (51), reposto em (56), abaixo:

56. [Chegamos à casa da minha melhor amiga,] [disse-lhe: [- Vamos embora] [estamos atrasadas]] [ela disse: [- Só são 12h da tarde] [vamos chegar lá às 15h30m.]] [- Ok vamos] [estou ansiosa [para começar as férias.]] [- Eu também] [disse ela.] (dsra_1_5h_mda)

Nesta tese, a elipse não é tida em conta, pelo que construções como [- Eu também] não são tratadas como unidades oracionais. Nas construções de comparação, como, por exemplo em [Depois deste dia ficamos mais amigos do que nunca] (gmp_1_10a_fdl), também não se faz qualquer operação de recuperação. Esta decisão teve por base unicamente uma questão de foco e exequibilidade de investigação.

Configuram-se como uma exceção à regra acima as construções em que está elidido apenas o verbo auxiliar, tendo sido já explicitado numa estrutura anterior, como se vê abaixo, em que é possível recuperar 'podem' – sinalizado por (-) – enquanto auxiliar do verbo pleno 'fazer':

57. [Podem fazer aquilo [a que chamam o amigo virtual,]] [ou então (-) fazer comentários terríveis para as pessoas nas redes sociais.]

Permanece a exceção quando o conetor oracional está elidido:

58. [[vê-lo através da web] é muito bom] [porque podes vê-lo] [e matar saudades] (ajfcb_2_7b_mda)

5. O segmento constituído por uma oração simples (sem qualquer oração a ela subordinada), independentemente de ela fazer parte de um composto de coordenação ou de suplementação ou ainda de ser apenas justaposta a outra, foi identificado como unidade 'simples'. Ao contrário disto, se o segmento contivesse uma oração subordinada (ou mais), foi anotado com o rótulo de 'subordinante'.

6. No caso de a oração ser bifuncional, se é subordinada de uma oração e subordinante de outra, prevaleceu a anotação como 'subordinada'. Se é coordenada com uma oração e subordinante de outra, prevaleceu a anotação como 'subordinante'. Vejam-se os exemplos:

59. [Agora do Messenger é mesma coisa] [também é muito perigoso] [como pede a morada de pessoa [que não conhecem.]] (raod_2_7b_mda)
60. [até que um amigo meu teve a ideia [de irmos dar um mergulho no mar,]] [de início todos achamos uma estupidez,] [mas depois pensamos melhor] [e vimos [que [se fôssemos] ia ser algo de marcante] [e que nos iríamos lembrar no futuro.]] (am1_1_10a_fdl)

Em (59), a oração [como pede a morada de pessoa que não conhecem.] é, simultaneamente, subordinada (causal) de [também é muito perigoso] e subordinante (ou, mais especificamente, contém um nome subordinante) de [que não conhecem.]. No exemplo (60), a oração [e vimos [que [se fôssemos] ia ser algo de marcante]], ao mesmo tempo que está coordenada com [mas depois pensamos melhor], é subordinante da completiva [que [se fôssemos] ia ser algo de marcante], que, por sua vez, está em coordenação com a oração [e que nos iríamos lembrar no futuro.].

9.7. Tipos e subtipos de oração subordinada

A classificação dos tipos de orações subordinadas⁶⁸ conforma-se às orientações de Mória (1996), Duarte (2003), Brito e Duarte (2003), Brito (2003), Brito e Matos (2003), Peres e Mascarenhas (2006), Perini (2010), Marques (2013), Mendes (2013), Barbosa (2013), Raposo (2013), Barbosa e Raposo (2013), Lobo (2013) e Veloso (2013), entre outros. São também referências Quirk *et al.* (1985), Greenbaum (1996), Huddleston e Pullum (2006, 2007), Schachter e Shopen (2007) e Biber *et al.* (2007). Destes estudos, chegou-se a quatro tipos básicos de orações subordinadas, abaixo apresentadas com exemplos que, a depender, satisfazem tanto casos de realização finita, como de não-finita:

i) Orações completivas

Funcionam como argumentos selecionados por um núcleo, que pode ser verbal, nominal ou adjetival, localizado tipicamente numa oração subordinante. A função sintática que a oração completiva estabelece em relação à oração subordinante pode ser de sujeito (ou subjetiva) (cf. (61)), de objeto direto (ou objetiva) (cf. (62)) ou oblíqua- (cf. (63)). Pode também ter a função de especificador do sujeito (cf. (64)). As orações completivas são conhecidas alternativamente por nominais (Biber *et al.* 2007), substantivas (Cunha e Cintra 2010; Bechara 2009) ou ainda de conteúdo (Huddleston e Pullum 2006, 2007). Vejam-se primeiramente exemplos de orações completivas que têm os seus núcleos verbais expressos numa forma finita, isto é, com o verbo no modo indicativo ou no modo conjuntivo:

61. Dois quarteirões mais adiante, é possível [que esteja alguém à beira do passeio]. (CPRC-noCOD_1039243)
62. Quando eu jogava nos Philadelphia 76ers, sabia [que a equipa não tinha talento absolutamente nenhum]. (CPRC-J64078)

⁶⁸ Os termos em negrito são os mesmos utilizados na construção do esquema de anotação do programa UAM Corpus Tool (UCT), apresentado ao fim deste anexo.

63. Os palestinianos insistem em [que não querem renegociar a questão de Hebron]. (CPRC-noCOD_1034536)

64. O problema é [que o vinho é mal pago em Portugal]. (CPRC-noCOD_1024739)

É importante salientar que a completiva no exemplo (64) se classifica na linha da tradição gramatical como uma oração predicativa de sujeito. No entanto, subscrevo o que diz Barbosa (2013: 1858), segundo a qual, em construções como a que se exemplifica em (64), “não é inteiramente adequado afirmar” que há uma relação de predicação entre o sujeito e a completiva, já que esta “não atribui nenhuma propriedade à entidade representada” por aquele, antes especifica-lhe o conteúdo, razão da nomenclatura “copulativa especificativa”, como prefere a autora, e também Raposo (2013: 1292).

A oração completiva finita pode funcionar ainda como complemento nominal (cf. (65)) ou adjetival (cf. (66)) localizado na oração subordinante:

65. “Começou a gritar, com medo [de que o homem lhe pudesse fazer mal novamente]”, concluiu a irmã da menina. (CPRC-noCOD_1026053)

66. Em relação ao futuro, os responsáveis portistas estão conscientes de [que só o tempo resolverá o problema da relva]. (CPRC-J106337)

As completivas também podem ser não-finitas, tendo o seu núcleo verbal no infinitivo, flexionado ou não-flexionado (às vezes precedido de preposição, caso a que Duarte (2003: 596) chama de infinitivo gerundivo), no gerúndio ou no particípio passado. As orações completivas não-finitas mantêm os mesmos tipos de relações sintáticas que as finitas, como se ilustra em (67), em que a oração infinitiva é sujeito; em (68), em que é objeto direto; em (69), em que é oblíqua; e em (70), em que é especificadora do sujeito. Podem as orações completivas não-finitas de infinitivo exercer ainda a função de predicativo do sujeito (cf. (71)), enquanto um dos tipos de construção de infinitivo preposicionado (Raposo 2013: 1294):

67. “É difícil [lembrar um período em que os indicadores fundamentais tenham sido tão bons]”, afirma o FMI. (CPRC-J34136)

68. Depois de conversar com Arizinho, Quitéria acredita [ter encontrado, finalmente, o seu filho]. (CPRC-J65303)

69. Jay-Z insiste em [apresentar-se não só como tendo concretizado esse sonho, mas igualmente como tendo escapado ao pesadelo do gueto]. (CPRC-J107666)
70. O problema é [saber o que acontece aos músicos do teatro de ópera quando desaparecer a sua entidade patronal, a empresa pública do Teatro Nacional de São Carlos]. (CPRC-J0038)
71. A Croácia jogava à sua maneira, muito lenta, com quase toda a gente [a dar duas voltas sobre si próprio e uma terceira em torno do companheiro mais próximo antes de dar a bola jogável]. (noCOD_1085037)

As orações completivas de nomes e de adjetivos também podem ser não-finitas, como se vê em (72) e (73), abaixo:

72. Tenho medo [de perder os olhos azuis que já ame], tenho medo de [ele se enterrar na sua própria cova]. (CPRC-J3453)
73. Ra's al Ghul vive obcecado [em conquistar o mundo], e tendo vivido mais de 600 anos, está ansioso por permanecer imortal. (CPRC-noCOD_1069748)

Nos exemplos (73) a (76), é possível notar que as orações completivas finitas são, em geral, introduzidas pelos complementadores *que* e *se*, mas, em (74), abaixo, em que a oração completiva funciona como objeto direto, mostra-se um caso específico em que tal tipo de orações, em particular se selecionado por verbos de inquirição (*pedir*, *perguntar*, etc.) e por certos verbos declarativos (*dizer*, *decidir*, *mostrar*, etc.) e epistémicos (*considerar*, *imaginar*, *descobrir*, etc.), também pode ser introduzido por complementadores interrogativos (*onde*, *como*, *quando*, etc.).

74. Pilatos perguntava [onde estava a verdade]. (CPRC-J58640)

Outra especificidade observa-se com as completivas de verbos declarativos de ordem (*dizer*, *pedir*, *solicitar*, etc.) que são introduzidas pelo item *para*. Neste caso, a preposição é alçada à condição de um complementador lexical (Duarte 2003: 621), como em (75):

75. ... só me apercebi bem da gravidade da situação quando a minha chefe disse [para fugirmos todos lá para fora]. (CPRC-noCOD_1069319)

É interessante notar, apenas a título de curiosidade, que Bechara (2009: 532-33) trata este tipo de construção como um caso de omissão do objeto direto do verbo *dizer*, sendo a oração dependente uma adverbial de finalidade. Ressalta o autor que “ainda hoje esta maneira condensada de dizer é repudiada, apesar da insistência com que penetra na linguagem das pessoas cultas”.

ii) Orações adverbiais

As orações adverbiais funcionam como adjuntos – não-argumentos – de outra oração, pelo que não são selecionadas pela oração a que se subordinam, funcionando como adjuntos de um predicado (ou frase). Podem ser finitas ou não-finitas, tal como acontece com outras orações subordinadas. Malgrado existir um conjunto de propriedades de natureza sintática que poderão caracterizar as orações adverbiais, como, por exemplo, a possibilidade de estar anteposta à oração subordinante, ou mesmo entre o sujeito e o predicado, estas orações são tipicamente classificadas pelos valores semânticos que veiculam em relação ao domínio proposicional, distinguindo-se os seguintes tipos principais: **condicionais**, **causais**, **finais**, **concessivas** e **temporais**.

As orações condicionais finitas (cf. (76), (77) e (78)) são tipicamente introduzidas pelo conector *se*, ou adverbializador, como preferem Schachter e Shopen (2007: 49), podendo a oração subordinante, neste caso, ser iniciada pelo advérbio *então* (cf. 78)). Outros itens conectores possíveis na oração dependente são: *caso*, *a menos que*, *a não ser que*, *desde que*, *contanto que*, etc.-

76. “[Se alteramos apenas pequenos aspetos], não conseguimos grandes resultados”. (CPRC-A185057)
77. [Caso se opte pela localização para sul], criaria, segundo o autarca, um “cluster” isolado. (CPRC-noCOD_1057169)
78. [Se houvesse consenso], então poderíamos talvez inverter a ordem destes dois pontos da primeira parte da ordem do dia. (CPRC-A146692)

A forma não-finita da oração condicional tanto pode ser introduzida pelas expressões conectivas *a* (seguida de infinitivo), *no caso de*, *na condição de*, etc. (cf. (79)), como pode não apresentar conector algum, estando o verbo numa forma nominal (cf. (80)):

79. [No caso de se registar no interior de um autocarro alguma ocorrência que faça perigar a segurança dos passageiros ou do motorista], este aciona um alarme que faz uma ligação direta, através da rede de telefones móveis, para o centro de controlo da empresa. (CPRC-J103884)
80. Desta forma, [não se concretizando a curto prazo o acordo interpartidário maioritário], deverá ter lugar a terceira alternativa. (CPRC-A124561)

Quanto às orações causais, os conectores mais comuns na realização finita das orações causais são *porque*, *pois*, *que*, *como*, *uma vez que*, *visto que*, *já que*, *dado que*. Como se depreende desta lista, assume-se, neste estudo, a noção de causa num sentido bastante lato, o que permite incluir, no mesmo grupo, não somente as orações de valor causal propriamente ditas (Lobo 2003: 14), mas também as orações de valor explicativo ou justificativo. As primeiras, conhecidas ainda como causa *de re* – ou de enunciado –, Lopes (2012: 453) define como construções que explicitam “um nexos de causa/consequência entre situações do mundo descritas ao nível dos conteúdos proposicionais articulados através do conector” (cf. (81) e (83), parafraseáveis por (82) e (84)); as segundas, a que se podem chamar também como causa *de dicto* – ou de enunciação –, Lopes (*idem*) define como construções em que “o nexos causal não é estabelecido com a predicação expressa na tradicionalmente chamada oração principal, mas sim com a enunciação dessa mesma predicação” (cf. (85) e (87), parafraseável por (86) e (88))⁶⁹:

⁶⁹ A aplicação de testes de parafraseamento ajudar a revelar a dificuldade em fazer a distinção entre causais e explicativas. O exemplo (84) pode ser parafraseável por “O facto de todo aquele litoral ser um autêntico pesqueiro de pargos, (...) leva-me a concluir que a costa de Aljezur é o paraíso dos pescadores”, indicando que a subordinada em (84) ([pois todo aquele litoral é um autêntico pesqueiro de pargos, sargos, congros, corvinas, douradas, besugos, robalos, entre outros peixes saborosos]) pode ser interpretada também como uma oração causal de enunciação (a par da leitura em (85)). Também o exemplo (88) permite duas leituras, podendo ser parafraseável por “O facto de o futebolista rescindir com a Real Sociedad antes do final do seu contrato e contra a vontade dos dirigentes do clube], (...) fez

81. De acordo com Madalena Sá, os trabalhadores aceitaram a proposta [porque a administração finlandesa garantiu o pagamento da totalidade das indemnizações. (CRPC-J104487)]
82. O facto de a administração finlandesa garantir o pagamento da totalidade das indemnizações fez com que os trabalhadores aceitassem a proposta.
83. A Costa de Aljezur é o paraíso dos pescadores desportivos, [pois todo aquele litoral é um autêntico pesqueiro de pargos, sargos, congros, corvinas, douradas, besugos, robalos, entre outros peixes saborosos]. (CRPC-J65904)
84. O facto de todo aquele litoral ser um autêntico pesqueiro de pargos, sargos, congros, corvinas, douradas, besugos, robalos, entre outros peixes saborosos faz com que a Costa de Aljezur seja o paraíso dos pescadores desportivos.
85. Escrever um romance contemporâneo é mais fácil, [porque há muitas coisas que já vimos na televisão, no dia a dia, que o cinema já mostrou]. (CPRC-noCOD_1048466)
86. O facto de haver muitas coisas que já vimos na televisão, no dia a dia, que o cinema já mostrou leva-me a concluir que escrever um romance contemporâneo é mais fácil.
87. [Visto que o futebolista rescindiu com a Real Sociedad antes do final do seu contrato e contra a vontade dos dirigentes do clube], o Valência terá de pagar o elevado montante determinado. (CPRC-J19696)
88. O facto de o futebolista ter rescindido com a Real Sociedad antes do final do seu contrato e contra a vontade dos dirigentes do clube leva-me a concluir que o Valência terá de pagar o elevado montante determinado.

Quanto à forma não-finita da oração causal, tanto pode ser introduzida pelas expressões conectivas *por causa de*, *devido ao facto de*, *visto*, *por*, *dado*, etc. (cf. (89)) seguidas

com que o Valência tenha de pagar o elevado montante determinado (...)", o que quer dizer que a oração [Visto que o futebolista rescindiu com a Real Sociedad antes do final do seu contrato e contra a vontade dos dirigentes do clube] também pode ser interpretada como uma oração causal de enunciado.

de verbo no infinitivo, como pode não apresentar conector algum, estando o verbo numa forma gerundiva (cf. (90)):

89. [Devido ao facto de levarmos uma vida pouco exposta, privada até em muitos sentidos,] as pessoas daqui e de fora são levadas a imaginar muito a nossa respeito. (CPRC-R1765)
90. [“Não tendo sido possível o acordo amigável com os expropriados,] a fixação da indemnização será feita de acordo com o disposto no código das expropriações, pelo que o processo correrá os seus trâmites nos termos legais”. (CPRC-noCOD_1056283)

As orações adverbiais finais (cf. (91) e (92)) são tipicamente introduzidas pelos conectores *para que*, *a fim de que*, etc.

91. Um trabalho tão enraizado localmente, como o das juntas de freguesia, deve ser divulgado junto das populações a quem se destina, [para que compreendam e apoiem a ação autárquica, praticamente voluntária]. (CPRC-J83735)
92. [A fim de que se conclua o processo iniciado em Oslo], é preciso que demonstrem uma vez mais a coragem de alterar posições de longa data, tal como o fizeram há sete anos. (CPRC-J61812)

A expressão não-finita das orações finais tipicamente é iniciada pela preposição *para* (cf. (93)), ou ainda pela locução conectiva *a fim de* (cf. (94)), ambas seguidas de infinitivo:

93. Em vez disso, [para poder manter o seu controlo sobre Gaza], Arafat foi obrigado a confiar nas pessoas que mais odiava. (CPRC-noCOD_1056338)
94. [A fim de inverter a tendência das nossas sociedades gangrenadas pelo desemprego], seria desejável que a União Europeia conseguisse criar 15 milhões de postos de trabalho até ao final do século. (CPRC-L0644)

As orações adverbiais de concessão finitas (cf. (95), (96) e (97)) expressam-se introduzidas pelos conectores *embora*, *conquanto*, *mesmo que*, *mesmo se*, *ainda que*, *posto que*, *(se) bem que*, *apesar de que*:

95. O Grande Prémio Bial é o maior prémio médico atribuído em Portugal, [ainda que esteja longe de ser – como foi ontem dito e repetido na cerimónia, erradamente – “o maior prémio na área da saúde depois do Nobel”]. (CPRC-noCOD_1077495)
96. [Mesmo que Monteiro queira reconsiderar a sua posição] terá que ser muito habilidoso, pois está preso às suas palavras. (CPRC-noCOD_1033768)
97. [Conquanto os fatores baixistas possam ter sido considerados de modo exagerado], a influência das taxas de juros tem estado a ser determinada pela orientação na circulação monetária. (CPRC-J1852)

A expressão não-finita das orações concessivas, além de se apresentar com o verbo no infinitivo (cf. (98)), também se apresenta frequentemente com o verbo quer numa forma participial (cf. (99)), quer no gerúndio (cf. (100)), acompanhado dos conectores *embora*, *mesmo*, *apesar de*, etc.

98. Sousa Lopes, [apesar de não ter conhecimento direto da situação], confirmou que essa desconfiança circulou no seio do PS local. (CPRC-J101714)
99. [Conquanto nascido em Caminha], Sidónio Pais tem ascendência em Barcelos. (CPRC-J3802)
100. O recurso ao anonimato, [embora fornecendo ao leitor informação que lhe permita situar a fonte], é uma prática demasiado vulgarizada no PÚBLICO. (CPRC-jpub_970921_oprov)

As orações temporais finitas (cf. (101)-(104)) são introduzidas pelos conectores *quando*, *enquanto*, *ao mesmo tempo que*, *antes que*, *depois que*, *assim que*, *logo que*, *desde que*, *até que*, etc.

101. O namorado conta que [quando lá foi buscar a pílula] acabou por trazer também preservativos. (CRPC-noCOD_1070200)
102. [Depois que acabou o seu curso de programador de computadores] nunca mais arranjou emprego. (CPRC-noCOD_1028850)

103. No edifício em frente, já fora da zona olímpica, lá está pendurado o pano com uns dizeres que ainda hei de decifrar [antes que os Jogos cheguem ao fim]. (CPRC-noCOD_1065735)

104. “... deixa-se cozer em lume brando [até que o molho fique bastante grosso e reduzido].” (CPRC-V0597)

As orações iniciadas por *até que* podem, por vezes, ser precedidas por pontuação forte, como ponto final ou ponto-e-vírgula. Nestes casos, Lobo (2013: 2001, nota 29), garante que “embora a oração temporal pareça funcionar como oração autónoma, na realidade, está implícito o contexto linguístico anterior”, como se vê em (105), abaixo:

105. Vão cantar o fado ao lado de nomes já consagrados como Vicente da Câmara, Nuno da Câmara Pereira, João Braga, Camané ou Florência. [Até que a voz lhes doa]. (CRPC-J19621)

Também se consideram como temporais finitas, ou ainda um seu subtipo, como prefere Lobo (2013: 2005), as orações iniciadas pelo conector *à medida que*, como em (106):

106. Os exércitos do Inimigo multiplicaram-se [à medida que os dos Homens diminuía em número e poder]. (CRPC-L0600)

A oração exemplificada em (106) expressa uma correlação escalar entre dois valores, estando um deles em relação de divergência com o outro, razão por que Marques (2013: 2164) nomeia tais construções de estruturas de correlação de escalas, cujo modelo são as construções com os operadores comparativos *quanto... mais*, *quanto... menos*, etc. (cf. seção iv, a seguir, que trata das orações gradativas). Assumo-as como temporais, e não as incluo no grupo das orações gradativas pelo evidente caráter temporal que se estabelece entre os domínios comparados. É como se vê nas palavras de Marques (2013: 2165): “O significado [deste tipo de orações] passa necessariamente pela consideração da escala do tempo, enquanto noutras construções de correlação de escalas [...], as escalas consideradas não estão necessariamente associadas a este domínio”.

As temporais não-finitas expressam-se com o verbo no infinitivo (cf. (107)) e podem ser introduzidas por *ao*, *até*, *depois de*, *antes de*, etc., com o verbo no gerúndio (cf. (108)), às vezes precedido da preposição *em*, ou com o verbo no particípio (cf. (109)):

107. Quanto a Washington, afirma que não fará concessões [antes de o programa nuclear norte-coreano, numa fase de desenvolvimento que ao certo se desconhece, comece a ser desmantelado sob supervisão internacional]. (CPRC-J108357)
108. Pedimos muitas desculpas se ofendemos alguém. [Tendo dito isso], temos que assumir a nossa quota de responsabilidade. (CPRC-J33793)
109. [Passados quatro anos], Portugal tem pela primeira vez uma Estratégia Nacional de Luta Contra a Droga. (CPRC-J95023)

Para além das orações causais, temporais, finais, concessivas e condicionais, acima apresentadas, Lobo (2001: 2027 e seguintes) refere ainda outros tipos de adverbiais que não costumam figurar em gramáticas prescritivas ou descritivas. São elas as orações de **circunstância negativa** (cf. (110)), iniciadas por *sem (que)*; de **comentário** (cf. (111)), iniciadas por *como*; **substitutivas** (cf. (112)), iniciadas por *em vez de*; e **acrescentativas** (cf. (113)), iniciadas (*para*) *além de*, as duas últimas com ocorrência tipicamente com verbos no infinitivo. Adiciona-se a estas orações o caso das orações de função exemplificativa (cf. (114)):

110. Colaram-lhe rótulos de social-democrata [sem que nunca tenha pertencido a qualquer partido]. (CPRC-J46803)
111. [Como era esperado], a cerimónia foi a mais concorrida de sempre. (CRPC-J21024)
112. Admita os seus próprios erros [em vez de culpabilizar os outros]. (CRPC-noCOD_1088730)
113. [Para além de subscrever todas as queixas que referi], queria fazer-lhe uma pergunta. (CRPC-noCOD_1088730)
114. É preciso chamar à atenção para a necessidade de proteger o meio-ambiente e adoptar comportamentos mais ecológicos, [como por exemplo separar o lixo"], diz o presidente da câmara, José Manuel Custódio. (CRPC-J101726)

iii) Orações relativas

As orações relativas funcionam ou como modificadores na estrutura do sintagma nominal (nome), em que se localiza o seu antecedente, pelo que são restritivas (cf. (115) e (116), ou como comentários de um sintagma nominal ou da frase matriz a que se ligam, sendo não-restritivas, portanto (cf. (117) e (118)). Os morfemas relativos disponíveis em português incluem as formas invariáveis *que*, *o que*, *quem*, *onde*, *como*, *quando*, *quanto* e as formas variáveis *o qual*, *a qual*, *os quais*, *as quais*, *cujo*, *cuja*, *cujos*, *cujas*, *quanto*, *quanta*, *quantos*, *quantas*, a que Schachter e Shopen (2007: 49) chamam de relativizadores:

115. Ainda não há uma equipa definida para o departamento, mas a Casa da Música teve de assegurar os projetos [que foram herdados da equipa anterior]. (CRPC-noCOD_1040962)
116. Em 1991, o estaleiro naval de Archangelsk, no Mar Ártico, local [onde trabalhara toda a vida], estava parado. (CRPC-J58255)
117. Ignorando os apelos da Comissão de Saúde, [que aconselhou o uso de mangas compridas e calças aos que passearem nas zonas de risco (parques e arredores dos rios)], os novaiorquinos preferiram encher o corpo e a casa de “gadgets” antimosquito. (CRPC-J35621)
118. Tenho sim um grande apreço por essa região, já que sou filho de um transmontano, [o que muito me honra]. (CRPC-noCOD_1035025)

Embora a relativa de frase, destacada em (118), possa alternativamente ser expressa por um nome – *situação*, *facto*, *coisa*, *pessoa*, etc., acompanhada da relativa (*facto* [*que muito me honra*]) –às vezes, acompanhado por demonstrativos (*facto este* [*que muito me honra*]), passando estas subordinadas a serem restritivas, considera-se, em linha com Móia (1996), que o constituinte relativo *o que*, em (55), é um único item mórfico, e não a combinação de um demonstrativo mais um relativo, como assume a tradição.

As orações relativas finitas tipicamente apresentam-se com antecedente expresso, como nos exemplos acima, mas também podem ocorrer com um antecedente não expresso (cf. Móia 1996), como se vê em (119), (120) e (121):

119. Nunca oculto [o que me vai na alma]. (CRPC-J31479)
120. O trajeto demora cerca de uma hora e é belíssimo, mas não se recomenda a [quem sofrer de vertigens]. (CRPC-noCOD_1070200)

121. Outro momento muito aplaudido foi [quando um congressista de Baucau se lembrou de destacar algumas das marcas dos sistemas democráticos, ou seja, a separação de poderes]. (CRPC-J40716)

Dos exemplos acima, é possível substituir, em (119), *o que* por *a coisa que*, em (120), *quem* por *a pessoa que* e, em (121), *quando* por *o momento em que*, sendo *coisa*, *pessoa* e *momento* os antecedentes. Por estas construções exercerem uma função sintática na oração subordinante, Quirk *et al.* (1985: 1056) classificam-nas como argumentais. Comportamento semelhante assume Bechara (2009: 468), que as trata como orações substantivas relativas. Móia (1996: 186), ao contrário destes autores, prefere mantê-las, com base num conjunto de critérios sintáticos e semânticos, no grupo das relativas modificadoras de um sintagma nominal com núcleo não expresso. Veloso (2013: 2118-9) corrobora a classificação de Móia, lembrando que “os traços semânticos do relativo utilizado nestes casos permitem reconstruir os traços correspondentes no nome não realizado foneticamente que funciona como antecedente”. Assim, feita a recuperação do antecedente no exemplo (120), reescrito em (122), abaixo, em que o termo recuperado está sublinhado, torna-se evidente que a oração relativa (*que sofrer de vertigens*) exerce o papel de modificador de um nome (*pessoa*):

122. O trajeto demora cerca de uma hora e é belíssimo, mas não se recomenda à pessoa [que sofrer de vertigens]. (CRPC-noCOD_1070200)

iv) Orações gradativas

Estas orações expressam informação associada a uma ou mais escalas relevantes no significado de nomes, verbos, advérbios ou adjetivos que são passíveis de gradação. Incluem-se, neste grupo, as orações comparativas (cf. (123) e (124)), as consecutivas (cf. (125) e (126)) e as proporcionais (cf. (127)). Tal classificação é retirada de Marques (2013), que trata destas orações como pertencentes ao grupo das construções de grau em português. Para a combinação da oração gradativa com a oração matriz, funcionam como itens conectivos: nas comparativas, *como* e *(do) que*, por vezes em correlação contínua ou descontínua com os operadores comparativos (Marques 2013: 2144) *mais* e *menos*, *tanto* e *tão* e as formas sintéticas de comparação *melhor*, *pior*, *maior* e *menor*; nas consecutivas, o item con-

juntivo prototípico é *que*, em correlação tipicamente descontínua com operadores consecutivos como *tal*, *tanto* e *tão*; nas proporcionais, a conexão dá-se pelo conjunto operador de correlação *quanto... mais...* e variações, etc.

123. Evelyn faz com Adam mais [do que se faz com uma estátua]. (CRPC-noCOD_1052528)

124. A situação das finanças públicas é má e as receitas públicas não estão a crescer tanto [como deviam]. (CRPC-J85741)

125. É preciso que a ferida seja mortal ou que o animal esteja tão cansado [que a sua captura seja iminente e certa]. (CRPC-A30113)

126. O meu filho mais velho, 14 anos, escondeu-se debaixo da cama, eles tiraram-no e dispararam tantos tiros sobre ele [que eu não conseguia reconhecê-lo]. (CRPC-J100429)

127. [Quanto mais me criticam por me comportar de maneira pouco feminina], mais me provocam. (CRPC-noCOD_1044461)

A classificação das orações gradativas enquanto um dos grandes tipos de subordinação, e não como um subtipo das adverbiais, não é habitual. Vários autores assumem as orações comparativas como um tipo de subordinada relativa, como, por exemplo, van Valin (2004), Huddleston e Pullum (2006) ou Marques (2013). Para Marques (2013: 2148), as comparativas são orações relativas sem um antecedente expreso e constituem *per se* um sintagma de grau. A tradição gramatical inclui as comparativas no grupo alargado das orações adverbiais (Cunha e Cintra 2010; Bechara 2009). Matos e Brito (2002) e Matos (2003) aproximam as comparativas das estruturas de coordenação, porque, segundo as autoras, se caracterizam de modo semelhante quanto à mobilidade dentro da estrutura frásica e também porque tanto os operadores de comparação como as conjunções coordenativas podem introduzir constituintes sintagmáticos e frásicos.

O estatuto sintático do segundo termo nas construções de comparação também é motivo de discórdia. Para alguns autores, pode o segundo termo da comparação tanto ser expreso por constituintes sintagmáticos, como por constituintes oracionais. Ao contrário disto, outros, como Huddleston e Pullum (2006: 212), defendem que as comparativas são

formas reduzidas da oração matriz, acontecendo a redução por meio da elipse de algum dos elementos expresso na oração anterior.

É importante referir também que o emprego da expressão *oração proporcional* segue a tradição gramatical brasileira, por ser óbvia e por demarcá-la mais fortemente dos outros tipos de estruturas gradativas. Estes tipos de oração não têm nomenclatura própria nas gramáticas portuguesas (às vezes, nem mesmo são referidas, como acontece no Dicionário Terminológico). São tratadas por Marques (2013: 2143) pelo rótulo funcional de *construções de correlação de escalas*.